FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Pancreatic-derived Factors, and Uses Related Thereto

Background of the Invention

Pattern formation is the activity by which embryonic cells form ordered spatial arrangements of differentiated tissues. The physical complexity of higher organisms arises during embryogenesis through the interplay of cell-intrinsic lineage and cell-extrinsic signaling. Inductive interactions are essential to embryonic patterning in vertebrate development from the earliest establishment of the body plan, to the patterning of the organ systems, to the generation of diverse cell types during tissue differentiation (Davidson, E., (1990) *Development* 108: 365-389; Gurdon, J. B., (1992) *Cell* 68: 185-199; Jessell, T. M. et al., (1992) *Cell* 68: 257-270). The effects of developmental cell interactions are varied. Typically, responding cells are diverted from one route of cell differentiation to another by inducing cells that differ from both the uninduced and induced states of the responding cells (inductions). Conversely, sometimes cells induce their neighbors to differentiate like themselves (homoiogenetic induction). Cell interactions in early development may be sequential, such that an initial induction between two cell types leads to a progressive amplification of diversity. Moreover, inductive interactions occur not only in embryos, but in adult cells as well, and can act to establish and maintain morphogenetic patterns as well as induce differentiation (J.B. Gurdon (1992) *Cell* 68:185-199).

Several classes of secreted polypeptides are known to mediate the cell-cell signaling that determines tissue fate during development. An important group of these signaling proteins are the TGFβ superfamily of molecules, which have wide range of functions in many different species. Members of the family are initially synthesized as larger precursor molecules with an amino-terminal signal sequence and a pro-domain of varying size (Kingsley, D.M. (1994) *Genes Dev.* 8:133-146). The precursor is then cleaved to release a mature carboxy-terminal segment of 110-140 amino acids. The active signaling moiety is comprised of hetero- or homodimers of the carboxy-terminal segment (Massague, J. (1990) *Annu. Rev. Cell Biol.* 6:597-641). The active form of the molecule then interacts with its receptor, which for this family of molecules is composed of two distantly related transmembrane serine/threonine kinases called type I and type II receptors (Massague, J. et al. (1992) *Cell* 69:1067-1070; Miyazono, K. A. et al. *EMBO J.* 10:1091-1101). TGFβ binds directly to the type II receptor, which then recruits the type I receptor and modifies it by phosphorylation. The type I receptor then transduces the

signal to downstream components, which are as yet unidentified (Wrana et al, (1994) *Nature* 370:341-347).

Several members of the TGFβ superfamily have been identified which play salient roles during vertebrate development. Dorsalin is expressed preferentially in the dorsal side of the developing chick neural tube (Basler et al. (1993)*Cell* 73:687-702). It promotes the outgrowth of neural crest cells and inhibits the formation of motor neuron cells *in vitro*, suggesting that it plays an important role in neural patterning along the dorsoventral axis. Certain of the bone morphogenetic proteins (BMPs) can induce the formation of ectopic bone and cartilage when implanted under the skin or into muscles (Wozney, J.M. et al. (1988) *Science* 242:1528-1534). In mice, mutations in BMP5 have been found to result in effects on many different skeletal elements, including reduced external ear size and decreased repair of bone fractures in adults (Kingsley (1994) *Genes Dev.* 8:133-146). Besides these effects on bone tissue, BMPs play other roles during normal development. For example, they are expressed in non skeletal tissues (Lyons et al. (1990) *Development* 109:833-844), and injections of BMP4 into developing *Xenopus* embryos promote the formation of ventral/posterior mesoderm (Dale et al (1992) *Development* 115:573-585). Furthermore, mice with mutations in BMP5 have an increased frequency of different soft tissue abnormalities in addition to the skeletal abnormalities described above (Green, M.C. (1958) *J. Exp. Zool.* 137:75-88). BMP2 and BMP7 have been found to be coexpressed in a number of tissues that are known to be the source of inductive signals, including the zone of polarizing activity and apical ectodermal ridge of the developing limb and the notochord, raising the possibility that BMP2/7 heterodimers may mediate aspects of these tissue interactions. BMP 2 transcripts have been found to be restricted within the developing gut to dorsal endoderm. (Lyons et al. (1995) *Mech Dev.* 50:71-83).

Another important family of signaling molecules is the wnt family. Wnt proteins are generally glycosylated, ranging in molecular weight from about 38-42 kDa. They also have features typically associated with secreted growth factors, including a hydrophobic signal peptide, the absence of additional transmembrane domains, highly conserved cysteine residues, and the presence of N-linked glycosylation sites (McMahon, A. (1992) *Trends Genet.* 8:91). Wnt proteins have been found to be associated with the cell surface and extracellular matrix (Papkoff, and Schryver (1990) *Mol. Cell. Biol.* 10:2723; Bradley, R., and Brown, A. (1990) *EMBO J.* 9:1569).

The wnt family of signaling molecules and potential proto-oncogenes are involved in carcinogenesis and in the regulation of pattern formation during embryogenesis and in the differentiation of cell lineages (Danielson et al. (1995). *J. Biol.*

Chem. 270:270:31225). Some of the wnt proteins also possess an ability to induce cell proliferation via an autocrine or paracrine route. (Danielson et al. *supra*). Increased levels of wnt expression have been associated with colorectal cancer (Vider et al. (1996) *Oncogene* 12:153) and mammary cancer (Bergstein et al. (1995) *Ann NY Acad Sci* 763:257; Lee et al. (1995) *Proc. Natl. Acad. Sci.* 92:2268). Wnt proteins have been shown to play a role in numerous developmental processes, including: the establishment of the dorsoventral axis (Cui et al. (1995) *Development* 121:2177), myogenesis (Munsterberg et al. (1995) *Genes & Development* 9:2911), and the induction of kidney epithelial morphogenesis (1994) *Developmental Biology* 166:815).

The function of embryonic endoderm during development is to construct the linings of two tubes within the body. The first tube, extending throughout the length of the body, is the digestive tube. Buds from this tube form the liver, gallbladder, and pancreas. The second tube, the respiratory tube, forms as an outgrowth of the digestive tube, and it eventually bifurcates into two lungs. Ectoderm produces three sets of cells during neurulation: neural tube cells, which give rise to neurons, glial cells, and ependyaml cells of the central nervous system; neural crest cells which give rise to peripheral nervous system cells, pigment cells, adrenal medulla, and certain areas of head cartilage; and the epidermis of the skin. Mesoderm forms connective tissue, blood cells, heart, the urogenital system, and parts of most of the internal organs (Gilbert, S.F. (1994) in *Developmental Biology* Sinaur Assoc.).

The term "liver" refers to the large, dark-red gland in the upper part of the abdomen on the right side, just beneath the diaphragm. Its manifold functions include storage and filtration of blood, conversion of sugars into glycogen, and many other metabolic activities. It also supplies bile to intestine. In adult vertebrates, this function is a minor one, but the liver originally arose as a digestive gland in lower chordates. Throughout the liver, a network of tiny tubules collects bile--a solution of salts, bilirubin (made when hemoglobin from red blood cells is broken down in liver), and fatty acids. Bile accumulates in the gall bladder, which empties into the small intestine by way of a duct. Bile has two functions in the intestine. First, it acts as a detergent, breaking fat into small globules that can be attacked by digestive enzymes. Second, and more important, bile salts aid in the absorption of lipids form the intestine; removal of the gall bladder sometimes causes difficulty with lipid absorption.

Digested food molecules absorbed into the bloodstream from the intestine pass directly to the liver by way of the hepatic portal vein. Before these molecules pass on into the rest of the body, the liver may change their concentration and even their chemical structure. The liver performs a vital role in detoxifying otherwise poisonous

substances. In addition, it stores food molecules that reach it from the intestine, converts them biochemically, and releases them back into the blood at a controlled rate. For instance, the liver removes glucose from the blood under the influence of the hormone insulin and stores it as glycogen. When the level of glucose in the blood falls, the hormone glucagon causes the liver to break down glycogen and release glucose into the blood.

The liver also synthesizes many of the blood proteins (e.g., albumins) and releases them into the blood when they are needed. In addition, the liver converts nitrogenous wastes into the form of urea for excretion by the kidneys. With the kidneys, the liver is vital in regulating what the blood contains when it reaches all the other organs of the body. Because the liver is the body's major organ for making all these biochemical adjustments, severe liver damage or loss of the liver is rapidly fatal.

The development of the eye is a complicated process. In Xenopus development the lens-forming ectoderm becomes competent to respond to lens-inducing signals from the presumptive neural plate during midgastrula stage. During late gastrula, the signals from neuralized ectoderm induces the presumptive lens-forming ectoderm. An additional inductive signal may be derived from the prospective foregut endoderm. At early neurula, the signals from the anterior neural region and endoderm have caused a lens-forming bias in the head ectoderm. At the late neurula stage, the optic vesicle contacts the lens-forming ectoderm, signaling the final determination of this tissue into lens. The optic cup induces the final determination of the lens. Mitogenic proteins in the anterior chamber maintain a line of proliferating cells in the ventral surface of the lens, while fibroblast growth factors stimulate the differentiation of dorsal lens epithelium. The corneal epithelium differentiates and secretes a primary stroma consisting of collagen layers. Endothelial cells secrete hyaluronic acid into the region, enabling mesenchymal cells from the neural crest to enter. Afterward, hyaluronidase digests the hyaluronic acid, causing the primary stroma to shrink. (Gilbert. 1994. *Developmental Biology*. Sinauer Associates, Inc.).

Like the eye, the mammalian kidney is an intricate structure. During embryogenesis, nephron formation is dependent on inductive interactions between cells of the metanephric blastema and ureteric bud, the two embryonic primordia that give rise to the kidney (Saxen (1987) " Organogenesis of the Kidney" Cambridge Univ. Press, Cambridge). Development of the kidney proceeds through three major stages. Early in development, the pronephric duct arises in the mesoderm just ventral to the anterior somites. The anterior region of the duct induces the adjacent mesenchyme to form the pronephric kidney tubules. The functional units of the metanephric kidney are the

nephrons, which are composed of glomerular and tubular segments lined by diverse epithelial cell types. The midportion of the nephric duct initiates the formation of the mesonephros. The second step of kidney development is the formation of the two ureter buds from the nephric ducts. The third stage of renal development consists of the interactions between ureter bud and metanephrogenic mesenchyme whereby some of the mesenchymal cells organize themselves into an epithelium. This is an example of reciprocal induction, since development signals derived from the ureteric bud induce metanephric mesenchymal cells to differentiate into nephron epithelia. In addition to the ureteric bud, a number of other tissues can act as heterologous inducers of this process in vitro, including embryonic spinal cord (Herzlinger et al. (1994) *Developmental Biology* 166:815). The wnt and TGFβ superfamily proteins have been found to be important in mediating nephron development.

The pancreas develops from the fusion of distinct dorsal and ventral diverticula (Gilbert, S.F. (1994) in *Developmental Biology* Sinaur Assoc. pp 361-362). As described below, in a preferred embodiment, the progenitor cells of the present invention are pancreatic or hepatic progenitor cells. The term "pancreas" is art recognized, and refers generally to a large, elongated, racemose gland situated transversely behind the stomach, between the spleen and duodenum. The pancreatic exocrine function, e.g., external secretion, provides a source of digestive enzymes. Indeed, "pancreatin" refers to a substance from the pancreas containing enzymes, principally amylase, protease, and lipase, which is used as a digestive aid. The exocrine portion is composed of several serous cells surrounding a lumen. These cells synthesize and secrete digestive enzymes such as trypsinogen, chymotrypsinogen, carboxypeptidase, ribonuclease, deoxyribonuclease, triacylglycerol lipase, phospholipase A_2, elastase, and amylase. Pancreatic islets develop from endodermal stem cells that lie in the fetal ductular pancreatic endothelium, which also contains pluripotent stem cells that develop into the exocrine pancreas. (Teitelman, G and J.K. Lee (1987) *Dev. Biol.* 121:454-466). Islet development proceeds through discrete developmental states during fetal gestation which are punctuated by dramatic transitions. The initial period is a protodifferentiated state which is characterized by the commitment of these pluripotent stem cells to the islet cell lineage, as manifested by the expression of insulin and glucagon. These protodifferentiated cells comprise a population of committed islet precursor cells which express only low levels of islet specific gene products and lack the cytodifferentiation of mature islet cells (Pictet, R. and W.J. Rutter (1972) Development of the embryonic endocrine pancreas. In *Endocrinology, Handbook of Physiology*, ed. R.O. Greep and E.B. Astwood, American Physiological Society: Washington D.C., p. 25-66). Islet formation involves an increase in islet specific gene expression as islets bud from the

pancreatic ducts; this process slows just before birth. Since differentiation of protodifferentiated precursors occurs during late fetal development of the pancreas, the factors regulating islet differentiation are likely to be expressed during this period (PCT/US93/12055).

The endocrine portion of the pancreas is composed of the islets of Langerhans. The islets of Langerhans appear as rounded clusters of cells embedded within the exocrine pancreas. Four different types of cells- α, β, δ, and ϕ-have been identified in the islets. The α cells constitute about 20% of the cells found in pancreatic islets and produce the hormone glucagon. Glucagon acts on several tissues to make energy available in the intervals between feeding. In the liver, glucagon causes breakdown of glycogen and promotes gluconeogenesis from amino acid precursors. The δ cells produce somatostatin which acts in the pancreas to inhibit glucagon release and to decrease pancreatic exocrine secretion. The hormone pancreatic polypeptide is produced in the ϕ cells. This hormone inhibits pancreatic exocrine secretion of bicarbonate and enzymes, causes relaxation of the gallbladder, and decreases bile secretion. The most abundant cell in the islets, constituting 60-80% of the cells, is the β cell, which produces insulin. Insulin is known to cause the storage of excess nutrients arising during and shortly after feeding. The major target organs for insulin are the liver, muscle, and fat-organs specialized for storage of energy.

Diabetes is characterized by pancreatic islet destruction or dysfunction leading to loss of glucose control. Diabetes mellitus is a metabolic disorder defined by the presence of chronically elevated levels of blood glucose (hyperglycemia). Insulin-dependent (Type 1) diabetes mellitus ("IDDM") results from an autoimmune-mediated destruction of the pancreatic β-cells with consequent loss of insulin production, which results in hyperglycemia. Type 1 diabetics require insulin replacement therapy to ensure survival. Non-insulin-dependent (Type 2) diabetes mellitus ("NIDDM") is initially characterized by hyperglycemia in the presence of higher-than-normal levels of plasma insulin (hyperinsulinemia). In Type 2 diabetes, tissue processes which control carbohydrate metabolism are believed to have decreased sensitivity to insulin. Progression of the Type 2 diabetic state is associated with increasing concentrations of blood glucose, and coupled with a relative decrease in the rate of glucose-induced insulin secretion. It is an extremely important disease with over 728,000 new cases of diabetes are diagnosed annually. With over 150,000 Americans dieing from the disease and its complications; the total yearly cost in the United States is over 20 billion dollars (Langer et al. (1993) *Science* 260:920-926).

The primary aim of treatment in both forms of diabetes mellitus is the same, namely, the reduction of blood glucose levels to as near normal as possible. Treatment of Type 1 diabetes involves administration of replacement doses of insulin. In contrast, treatment of Type 2 diabetes frequently does not require administration of insulin. For example, initial therapy of Type 2 diabetes may be based on diet and lifestyle changes augmented by therapy with oral hypoglycemic agents such as sulfonylurea. Insulin therapy may be required, however, especially in the later stages of the disease, to produce control of hyperglycemia in an attempt to minimize complications of the disease. The identification of factors which control the development and maintenance of pancreatic tissue could provide a cure for this illness.

Summary of the Invention

The present invention relates to the discovery of novel genes and their gene products, expressed in vertebrate organisms, referred to hereinafter as pancreatic derived factor genes or *PDF* genes, *PDF1* and *PDF2*.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

Figure 1 is a sequence comparison of a portion of the *PDF1* polypeptide with BMP-2A and BMP-2B/BMP-4 core sequence.

Detailed Description of the Invention

Of particular importance in the development and maintenance of tissue in vertebrate animals is a type of extracellular communication called induction, which occurs between neighboring cell layers and tissues (Saxen et al. (1989) *Int J Dev Biol* 33:21-48; and Gurdon et al. (1987) *Development* 99:285-306). In inductive interactions, chemical signals secreted by one cell population influence the developmental fate of a second cell population. Typically, cells responding to the inductive signals are diverted from one cell fate to another. Inductive signals are transmitted by key regulatory proteins that function during development to determine tissue patterning, as well as in adult tissue to maintain tissue integrity. For example, signals mediated by the TGFβ and wnt families have been shown to play a variety of roles, including participating in tissue induction and maintenance.

The present invention concerns the discovery of two novel gene products, referred to herein as "pancreatic derived factors" or "*PDFs*". *PDF1* has homology to members of the TGFβ-superfamily, in particular BMP2a (Schlunegger, M.P. and M.G. Grutter (1992) *Nature* 358:430-434). *PDF2* has homology to members of the wnt family. The *PDFs* appear to play a role in determining tissue fate and in maintenance of differentiated states. For instance, the results provided below indicate that proteins encoded by the *PDF* genes may participate in the control of development and maintenance of a variety of embryonic and adult tissues. For example, during embryonic induction, certain of the *PDFs* are implicated in the differentiation and patterning of endodermal tissue, and more particularly, in pancreatic, kidney, and eye development.

The cDNAs corresponding to *PDF* gene transcripts were initially cloned from pooled dorsal and ventral pancreatic buds isolated from e13- and e14-day mouse fetuses. Degenerate primers derived from BMP were used to amplify *PDF1* sequences and degenerate primers derived from Wnt were used to amplify *PDF2* sequences. Clones were picked randomly and one member of each class was sequenced completely. A partial nucleic acid sequence for *PDF1* is provided in SEQ. ID NO. 1, and the corresponding amino acid sequence is provided in SEQ. ID NO. 2. The full length

nucleic acid sequence of *PDF1* is provided in SEQ. ID NO. 5, and the corresponding amino acid sequence is provided in SEQ. ID NO. 6. The nucleic acid sequence for *PDF2* is provided in SEQ. ID NO. 3, and the corresponding amino acid sequence is provided in SEQ. ID NO. 4.

Analysis of *PDF1* sequences suggests approximately 50% amino acid sequence identity with the core sequence of TGFβ superfamily of signaling proteins, in particular BMP-2a, a member of the Vg/dpp subgroup. So far all isolated or predicted proteins of the TGF superfamily show sequence invariance in the cysteines 15, 44, 48, 77, 78, 109, and 111, as well as Pro 36 and Gly 46 (Schlunegger, M.P. and Grutter, M.G. (1992) *Nature* 358:430-434).

Careful inspection of the *PDF1* gene clone suggests that, as shown in the alignment presented in Figure 1, the clone possesses the critical cysteine and proline residues of the core TGFβ superfamily sequence.

The expression of these genes in the developing pancreas indicates a role for this novel gene product in the development of endodermal tissue and in particular in pancreatic development. Experimental evidence indicates a functional role for the *PDF1* in signal transduction mediated by members of the TGFβ superfamily.

Experimental evidence indicates a functional role for the *PDF2* in signal transduction mediated by members of the wnt family.

Accordingly, certain aspects of the present invention relate to nucleic acids encoding *PDF* proteins, the *PDF* proteins themselves, antibodies immunoreactive with *PDF* proteins, and preparations of such compositions. In addition, nucleic acids which encode fragments or the protein fragments themselves are also contemplated. Moreover, the present invention provides diagnostic and therapeutic assays and reagents for detecting and treating disorders involving, for example, aberrant expression (or loss thereof) of *PDF* homologs. In addition, drug discovery assays are provided for identifying agents which can modulate the biological function of *PDF* proteins, such as by altering the binding of *PDF* molecules to bind to their cognate receptors and thus modulate signaling via *PDF* signal transduction pathway. Such agents can be useful therapeutically to alter the growth and/or differentiation of a cell. Other aspects of the invention are described below or will be apparent to those skilled in the art in light of the present disclosure.

Another aspect of the present invention relates to the use of pancreatic-derived factor (PDF) proteins in the manufacture of a medicament for the treatment of a

pancreatic disorder or to modulate growth and/or differentiation of pancreatic cells or stem cells capable of differentiating to pancreatic cells.

For convenience, certain terms employed in the specification, examples, and appended claims are collected here.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

The terms peptide, polypeptide, and protein are used interchangeably herein.

The term "*PDF* polypeptide" refers to a family of polypeptides characterized at least in part by being identical or sharing a degree of sequence homology with all or a portion of the polypeptide represented in SEQ ID No. 2, 4 or 6. The *PDF* polypeptides can be cloned or purified from any of a number of eukaryotic organisms, especially vertebrates, and particularly mammals. Moreover, other *PDF* polypeptides can be generated according to the present invention, which polypeptides do not ordinarily exist in nature, but rather are generated by non-natural mutagenic techniques.

As described in the appended examples, a number of features of this family of proteins can be observed from comparison of a PDF polypeptide with other members of the TGFβ or wnt family of ligands. For example, with regard to *PDF1*, we have noted that a core sequence motif, consisting of three cysteine residues and one proline residue, is apparently conserved with approximately the same characteristic spacing within the primary sequence of each of the known TGFβ superfamily polypeptides. This conserved "motif" containing these three cysteines and one proline, referred to herein as a "core motif", may represent a fragment which retains certain biological activities of the full length (mature) protein, such as, for example, the ability to bind a TGFβ superfamily receptor. In exemplary *PDF1* polypeptides, the core motif is represented by residues 293-392 of SEQ ID No. 2.

Those skilled in the art, in light of the present invention, will be able to easily ascertain the equivalent core motifs in other *PDF* polypeptides. The core sequence motif presumably represents a fragment which is likely to be essential for the activity of the mature protein, with the exception of a secretion signal sequence and a glycosyl attachment sequence.

Finally, a "mature" *PDF* polypeptide refers to a PDF polypeptide which lacks a signal sequence (e.g., a peptidyl portion which causes extracellular secretion of at least a portion of the protein).

A "glycosylated" *PDF* polypeptide is an *PDF* polypeptide having a covalent linkage with a glycosyl group (e.g. a derivatized with a carbohydrate). For instance, the exemplary *PDF* and proteins contain one potential N-linked glycosylation site. To generate an unglycosylated *PDF* polypeptide, the polypeptide can be expressed in a system which is defective for glycosylation, such as a bacterial cell. Alternatively, an existing glycosylation site, such as ASN_{59} of *PDF* or the equivalent, can be mutated to preclude carbohydrate attachment. Likewise, new glycosylation sites, such as for N-linked or O-linked glycosylation, can be added by recombinant techniques.

The phrases "modifies or "modulates" cellular activities, with respect to the biological activity of the subject *PDF* polypeptides, refer to statistically significant changes which occur in a cell due to activation of intracellular signals, either primary or secondary, by *PDF* interaction with other cellular proteins. For example, such cellular activities which may be affected by *PDF* include proliferation, differentiation or survival of a cell, as well as cell-cell adhesion and other alterations in phenotype. In one aspect, the cellular activities which can be modified by a PDF polypeptide pertain to the development and maintenance of pancreatic tissue, in particular islet tissue. Other exemplary tissues which may be modified by a PDF polypeptide include, for example, kidney tissue and eye tissue. In general, the cellular modifications can be the relatively-direct biochemical consequence of signal transduction events or can be caused more indirectly, such as *PDF* dependent activation or inactivation of particular genes or gene programs. A *PDF* polypeptide which "modifies" cellular activities can refer to homologs which either mimic or inhibit the normal response of a cell to the wild-type form of the protein.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a PDF polypeptide of the present invention, including both exon and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding a PDF polypeptide and comprising *PDF*-encoding exon sequences, though it may optionally include intron sequences which are either derived from a chromosomal *PDF* gene or from an unrelated chromosomal gene. An exemplary recombinant gene encoding the subject *PDF* polypeptide is represented in the appended Sequence Listing. The term "intron" refers to a DNA sequence present in a given *PDF* gene which is not translated into protein and is generally found between exons.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transduction", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transduced cell expresses a recombinant form of a PDF polypeptide or, where anti-sense expression occurs from the transferred gene, the expression of a naturally-occurring form of the *PDF* protein is disrupted.

As used herein, the term "specifically hybridizes" refers to the ability of the probe/primer of the invention to hybridize to at least 15 consecutive nucleotides of a PDF gene, such as a PDF nucleic acid sequence designated in SEQ ID No: 1, 3 or 5, or a sequence complementary thereto, or naturally occurring mutants thereof, such that it has less than 15%, preferably less than 10%, and more preferably less than 5% background hybridization to a cellular nucleic acid (e.g., mRNA or genomic DNA) encoding a protein other than a PDF protein, as defined herein.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of the recombinant *PDF* gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are

different from those sequences which control transcription of the naturally-occurring forms of *PDF* proteins.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of hepatic or pancreatic origin, e.g. neuronal cells. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

As used herein, a "transgenic animal" is any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form the *PDF* protein, e.g. either agonistic or antagonistic forms. However, transgenic animals in which the recombinant *PDF* gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more *PDF* genes is caused by human intervention, including both recombination and antisense techniques.

The "non-human animals" of the invention include vertebrates such as rodents, non-human primates, sheep, dog, cow, chickens, amphibians, reptiles, etc. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse, though transgenic amphibians, such as members of the *Xenopus* genus, and transgenic chickens can also provide important tools for understanding and identifying agents which can affect, for example, embryogenesis and tissue formation. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that the recombinant *PDF* gene is present and/or expressed or disrupted in some tissues but not others.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., a PDF polypeptide, or pending an antisense transcript thereto), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding a PDF polypeptide" may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individual organisms, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

"Homology" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40 percent identity, though preferably less than 25 percent identity, with a PDF gene sequence of the present invention.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding a PDF polypeptide with a second amino acid sequence defining a domain (e.g. polypeptide portion) foreign to and not substantially homologous with any domain of the *PDF* protein. A chimeric protein may present a foreign domain which is found (albeit in a different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion of protein structures expressed by different kinds of organisms. In general, a fusion protein can be represented by the general formula X-*PDF*-Y, wherein *PDF* represents a portion of the fusion protein which is derived from a PDF protein, and X and Y are independently absent or represent amino acid sequences which are not related to the *PDF* sequence in an organism, including naturally occurring mutants.

As used herein, the terms "transforming growth factor-beta" and "TGFβ" denote a family of structurally related paracrine polypeptides found ubiquitously in vertebrates, and prototypic of a large family of metazoan growth, differentiation, and morphogenesis factors (see, for review, Massaque et al. (1990) *Ann Rev Cell Biol* 6:597-641; Massaque et al. (1994) *Trends Cell Biol.* 4:172-178; Kingsley (1994) *Gene Dev.* 8:133-146; and Sporn et al. (1992) *J Cell Biol* 119:1017-1021). As described in Kingsley, *supra*, the TGFβ superfamily has at least 25 members, and can be grouped into distinct sub-families with highly related sequences. The most obvious sub-families include the following: the TGFβ sub-family, which comprises at least four genes that are much more similar to TGFβ-1 than to other members of the TGFβ superfamily; the activin sub-family, comprising homo- or hetero-dimers or two sub-units, inhibinβ-A and inhibinβ-B. The decapentaplegic sub-family, which includes the mammalian factors BMP2 and BMP4, which can induce the formation of ectopic bone and cartilage when implanted under the skin or into muscles. The 60A sub-family, which includes a number of mammalian homologs, with osteoinductive activity, including BMP5-8. Other members of the TGFβ superfamily include the gross differentiation factor 1 (GDF-1), GDF-3/VGR-2, dorsalin, nodal, mullerian-inhibiting substance (MIS), and glial-derived neurotrophic growth factor (GDNF). It is noted that the DPP and 60A sub-families are related more closely to one another than to other members of the TGFβ superfamily, and have often been grouped together as part of a larger collection of molecules called DVR (<u>d</u>pp and <u>v</u>g1 <u>r</u>elated). Unless evidenced from the context in which it is used, the term TGFβ as used throughout this specification will be understood to generally refer to members of the TGFβ superfamily. Reference to specific members of the TGFβ sub-family will be explicit, or evidenced from the context in which the term TGFβ is used.

As used herein the term "wnt" is meant to refer to the wnt family of signaling molecules. Wnt proteins have been shown to play a role in numerous developmental processes, including: the establishment of the dorsoventral axis (Cui et al. (1995) *Development* 121:2177), myogenesis (Munsterberg et al. (1995) *Genes & Development* 9:2911), and the induction of kidney epithelial morphogenesis (1994) *Developmental Biology* 166:815).

The term "frizzled receptor" refers to the gene products of the tissue polarity gene Frizzled (Fz) and other structurally related receptors, e.g., which interact with (bind) Wnt proteins.

The term "isolated" as also used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, an isolated nucleic acid encoding one of the subject *PDF* polypeptides preferably includes no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks the *PDF* gene in genomic DNA, more preferably no more than 5kb of such naturally occurring flanking sequences, and most preferably less than 1.5kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state.

As described below, one aspect of the invention pertains to isolated nucleic acids comprising nucleotide sequences encoding *PDF* polypeptides, and/or equivalents of such nucleic acids. The term nucleic acid as used herein is intended to include fragments as equivalents. The term equivalent is understood to include nucleotide sequences encoding functionally equivalent *PDF* polypeptides or functionally equivalent peptides having an activity of a PDF protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the *PDF* cDNA sequence shown in SEQ ID No: 1, 3 or 5, due to the degeneracy of the genetic code. Equivalents will also include nucleotide sequences that hybridize under stringent conditions (i.e., equivalent to about 20-27°C below the melting temperature (T_m) of the DNA duplex formed in about 1M salt) to the nucleotide sequences represented in SEQ ID No: 1, 3 or 5. In one

embodiment, equivalents will further include nucleic acid sequences derived from and evolutionarily related to, a nucleotide sequence shown in SEQ ID No: 1, 3 or 5.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of the subject *PDF* polypeptide which function in a limited capacity as one of either a PDF agonist (mimetic) or a PDF antagonist, in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function, and with fewer side effects relative to treatment with agonists or antagonists which are directed to all of the biological activities of the naturally occurring form of *PDF* protein.

Homologs of the subject *PDF* protein can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For instance, mutation can give rise to homologs which retain substantially the same, or merely a subset, of the biological activity of the *PDF* polypeptide from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to the cognate receptor of the *PDF* protein, thus blocking signal transduction. In addition, agonistic forms of the protein may be generated which are constitutively active. Thus, the *PDF* protein and homologs thereof provided by the subject invention may be either positive or negative regulators of signal transduction via a TGFβ superfamily pathway.

In general, polypeptides referred to herein as having an activity (e.g., are "bioactive") of a PDF protein are defined as polypeptides which include an amino acid sequence corresponding (e.g., identical or homologous) to all or a portion of the amino acid sequence of the *PDF* protein shown in SEQ ID No: 2, 4 or 6 and which mimic or antagonize all or a portion of the biological/biochemical activities of a naturally occurring *PDF* protein. Examples of such biological activity include the ability to induce (or otherwise modulate) formation and differentiation of endodermal tissue of developing vertebrate embryos. The subject polypeptides can be characterized, therefore, by an ability to induce and/or maintain differentiation or survival of stem cells or germ cells, including cells derived from the primitive gut including; trachea, bronchi, lungs, digestive tube, pancreas, liver, allantois, or urinary bladder, and the like. Other stem cells which can be influenced by a PDF protein are derived from embryonic mesoderm, including chordamesoderm, dorsal mesoderm, intermediate mesoderm (including cells derived from the metanephric diverticulum, mesonephric ducts, mesonephros, metanephros, and the like), lateral mesoderm, or head mesenchyme. Still other stem cells which can be influenced by a PDF protein include cells derived from

embryonic ectoderm, including cells derived from the eye, the outer epithelium of the body, the neural tube, or the neural crest. *PDF* proteins of the present invention can also have biological activities which include an ability to regulate organogenesis. Additional effects of *PDF* may be seen on tissue maintenance and repair post-development.

The bioactivity of the subject *PDF* proteins may also include the ability to alter the transcriptional rate of a gene as, for example, a downstream component of a signal transduction cascade initiated by the interaction of a subject *PDF* protein with its cognate receptor.

Other biological activities of the subject *PDF* proteins are described herein or will be reasonably apparent to those skilled in the art. According to the present invention, a polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of a *PDF* protein.

Preferred polypeptides comprise an amino acid sequence at least 60% homologous, more preferably 70% homologous and most preferably 80% homologous with the amino acid sequence of *PDF*, e.g., such as SEQ ID No: 2, 4 or 6. Polypeptides at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology with an amino acid sequence represented in SEQ ID No: 1, 3 or 5 are of course also within the scope of the invention. In one embodiment, a PDF peptide has at least one activity of the subject *PDF* polypeptide.

In addition, fragments of a PDF polypeptides are provided by the present invention. Exemplary fragments of *PDF1* comprise the amino acid sequence ERQGNYCKKTPLYIDFKEIG or CKKTPLYIDFKEIGWD with fragments with comprise the amino acid sequence CKKTPLTIDFKEIG being more preferred. Other fragments, e.g., derived from or otherwise including at least a portion of the core sequence are also contemplated.

In other embodiments, preferred nucleic acids encode a bioactive fragment of a *PDF* polypeptide comprising an amino acid sequence at least 60% homologous, more preferably 70% homologous and most preferably 80% homologous with an amino acid sequence of SEQ ID No: 2, 4 or 6. Nucleic acids which encode polypeptides at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homologous, or identical, with an amino acid sequence represented in SEQ ID No: 2, 4 or 6 are also within the scope of the invention. In the case of *PDF1*, preferred fragments comprise the nucleic acids which encode the amino acids CKKTPLTIDFKEIG.

Another aspect of the invention provides a nucleic acid which hybridizes under high or low stringency conditions to the nucleic acid represented by SEQ ID No: 1, 3 or 5. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C.

Nucleic acids having a sequence that differs from the nucleotide sequences shown in SEQ ID No: 1, 3 or 5 due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent peptides (i.e., a peptide having a biological activity of a *PDF* polypeptide) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC each encode histidine) may result in "silent" mutations which do not affect the amino acid sequence of a *PDF* polypeptide. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject *PDF* polypeptides will exist among vertebrates. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about 3-5% of the nucleotides) of the nucleic acids encoding polypeptides having an activity of a *PDF* polypeptide may exist among individuals of a given species due to natural allelic variation.

As used herein, a PDF gene fragment refers to a nucleic acid having fewer nucleotides than the nucleotide sequence encoding the entire mature form of a *PDF* protein yet which (preferably) encodes a polypeptide which retains some biological activity of the full length protein. Fragment sizes contemplated by the present invention include, for example, 5, 10, 15, 25, 50, 75, 100, or 200 (contiguous) amino acids in length.

As indicated by the examples set out below, *PDF* protein-encoding nucleic acids can be obtained from mRNA present in any of a number of eukaryotic cells. It should also be possible to obtain nucleic acids encoding *PDF* polypeptides of the present invention from genomic DNA from both adults and embryos. For example, a gene encoding a *PDF* protein can be cloned from either a cDNA or a genomic library in accordance with protocols described herein, as well as those generally known to persons

skilled in the art. A cDNA encoding a *PDF* protein can be obtained by isolating total mRNA from a cell, e.g. a mammalian cell, e.g. a human cell, including embryonic cells. Double stranded cDNAs can then be prepared from the total mRNA, and subsequently inserted into a suitable plasmid or bacteriophage vector using any one of a number of known techniques. The gene encoding a *PDF* protein can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acid of the invention can be DNA or RNA. A preferred nucleic acid is a cDNA represented by a sequence shown in any of SEQ ID Nos:1, 3 or 5.

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. binds) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding one or more of the subject *PDF* proteins so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a *PDF* protein. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences of a *PDF* gene. Such oligonucleotide probes are preferably modified oligonucleotides which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and are therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidite, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by Van der Krol et al. (1988) *Biotechniques* 6:958-976; and Stein et al. (1988) *Cancer Res* 48:2659-2668.

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized

in a manner appropriate for antisense therapy in general. For such therapy, the oligomers of the invention can be formulated for a variety of loads of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous for injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which they specifically bind. Such diagnostic tests are described in further detail below.

Likewise, the antisense constructs of the present invention, by antagonizing the normal biological activity of *PDF*, can be used in the manipulation of tissue, e.g. tissue differentiation, both *in vivo* and in *ex vivo* tissue cultures.

This invention also provides expression vectors containing a nucleic acid encoding a PDF polypeptide, operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of the subject *PDF* proteins. Accordingly, the term transcriptional regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel; *Gene Expression Technology*: *Methods in*

Enzymology 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences-sequences that control the expression of a DNA sequence when operatively linked to it may be used in these vectors to express DNA sequences encoding the *PDF* polypeptides of this invention. Such useful expression control sequences, include, for example, a viral LTR, such as the LTR of the Moloney murine leukemia virus, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage λ , the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered. In one embodiment, the expression vector includes a recombinant gene encoding a peptide having an agonistic activity of a subject *PDF* polypeptide, or alternatively, encoding a peptide which is an antagonistic form of the *PDF* protein. Such expression vectors can be used to transfect cells and thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

Moreover, the gene constructs of the present invention can also be used as a part of a gene therapy protocol to deliver nucleic acids encoding either an agonistic or antagonistic form of the subject *PDF* protein. Thus, another aspect of the invention features expression vectors for *in vivo* transfection and expression of a PDF polypeptide in particular cell types so as to reconstitute the function of, or alternatively, abrogate the function of *PDF* in a tissue in which *PDF* is misexpressed; or to deliver a form of the protein which alters differentiation of tissue, or which inhibits neoplastic transformation, by modulating the biological function of a *PDF* receptors (e.g., the wnt family or TGFβ superfamily ligand receptors).

Expression constructs of the subject *PDF* polypeptide, and mutants thereof, may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the *PDF* gene to cells *in vivo*. Approaches include insertion of the subject gene in viral vectors including recombinant retroviruses,

adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or $CaPO_4$ precipitation carried out *in vivo*. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene therapy, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically. Furthermore, it will be recognized that the particular gene construct provided for *in vivo* transduction of *PDF* expression are also useful for *in vitro* transduction of cells, such as for use in the ex vivo tissue culture systems described below.

A preferred approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA, encoding the particular form of the *PDF* polypeptide desired. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells which have taken up viral vector nucleic acid.

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host. A major prerequisite for the use of retroviruses is to ensure the safety of their use, particularly with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed "packaging cells") which produce only replication-defective retroviruses has increased the utility of retroviruses for gene therapy, and defective retroviruses are well characterized for use in gene transfer for gene therapy purposes (for a review see Miller, A.D. (1990) *Blood* 76:271). Thus, recombinant retrovirus can be constructed in which part of the retroviral coding sequence (*gag, pol, env*) has been replaced by nucleic acid encoding one of the subject receptors rendering the retrovirus replication defective. The replication defective retrovirus is then packaged into virions which can be used to infect a target cell through the use of a helper virus by standard techniques. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* or *in vivo* with such viruses can be found in <u>Current Protocols in Molecular Biology</u>, Ausubel, F.M. et al. (eds.) Greene Publishing Associates, (1989), Sections 9.10-9.14 and other standard laboratory manuals. Examples

of suitable retroviruses include pLJ, pZIP, pWE and pEM which are well known to those skilled in the art. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include ψCrip, ψCre, ψ2 and ψAm. Retroviruses have been used to introduce a variety of genes into many different cell types *in vitro* and/or *in vivo* (see for example Eglitis, et al. (1985) *Science* 230:1395-1398; Danos and Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464; Wilson et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:3014-3018; Armentano et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6141-6145; Huber et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8039-8043; Ferry et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8377-8381; Chowdhury et al. (1991) *Science* 254:1802-1805; van Beusechem et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:7640-7644; Kay et al. (1992) *Human Gene Therapy* 3:641-647; Dai et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:10892-10895; Hwu et al. (1993) *J. Immunol.* 150:4104-4115; U.S. Patent No. 4,868,116; U.S. Patent No. 4,980,286; PCT Application WO 89/07136; PCT Application WO 89/02468; PCT Application WO 89/05345; and PCT Application WO 92/07573).

Furthermore, it has been shown that it is possible to limit the infection spectrum of retroviruses and consequently of retroviral-based vectors, by modifying the viral packaging proteins on the surface of the viral particle (see, for example PCT publications WO93/25234 and WO94/06920). For instance, strategies for the modification of the infection spectrum of retroviral vectors include: coupling antibodies specific for cell surface antigens to the viral *env* protein (Roux et al. (1989) *PNAS* 86:9079-9083; Julan et al. (1992) *J. Gen Virol* 73:3251-3255; and Goud et al. (1983) *Virology* 163:251-254); or coupling cell surface receptor ligands to the viral *env* proteins (Neda et al. (1991) *J Biol Chem* 266:14143-14146). Coupling can be in the form of the chemical cross-linking with a protein or other variety (e.g. lactose to convert the *env* protein to an asialoglycoprotein), as well as by generating fusion proteins (e.g. single-chain antibody/*env* fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, can also be used to convert an ecotropic vector in to an amphotropic vector.

Moreover, use of retroviral gene delivery can be further enhanced by the use of tissue- or cell-specific transcriptional regulatory sequences which control expression of the *PDF* gene of the retroviral vector.

Another viral gene delivery system useful in the present invention utilitizes adenovirus-derived vectors. The genome of an adenovirus can be manipulated such that it encodes and expresses a gene product of interest but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle. See for example Berkner et al.

(1988) *BioTechniques* 6:616; Rosenfeld et al. (1991) *Science* 252:431-434; and Rosenfeld et al. (1992) *Cell* 68:143-155. Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are well known to those skilled in the art. Recombinant adenoviruses can be advantageous in certain circumstances in that they are not capable of infecting nondividing cells and can be used to infect a wide variety of cell types (Rosenfeld et al. (1992) cited *supra*). Furthermore, the virus particle is relatively stable and amenable to purification and concentration, and as above, can be modified so as to affect the spectrum of infectivity. Additionally, introduced adenoviral DNA (and foreign DNA contained therein) is not integrated into the genome of a host cell but remains episomal, thereby avoiding potential problems that can occur as a result of insertional mutagenesis *in situ*ations where introduced DNA becomes integrated into the host genome (e.g., retroviral DNA). Moreover, the carrying capacity of the adenoviral genome for foreign DNA is large (up to 8 kilobases) relative to other gene delivery vectors (Berkner et al. cited *supra*; Haj-Ahmand and Graham (1986) *J. Virol.* 57:267). Most replication-defective adenoviral vectors currently in use and therefore favored by the present invention are deleted for all or parts of the viral E1 and E3 genes but retain as much as 80% of the adenoviral genetic material (see, e.g., Jones et al. (1979) *Cell* 16:683; Berkner et al., *supra*; and Graham et al. in Methods in Molecular Biology, E.J. Murray, Ed. (Humana, Clifton, NJ, 1991) vol. 7. pp. 109-127). Expression of the inserted *PDF* gene can be under control of, for example, the E1A promoter, the major late promoter (MLP) and associated leader sequences, the E3 promoter, or exogenously added promoter sequences.

Yet another viral vector system useful for delivery of the subject *PDF* gene is the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle. (For a review see Muzyczka et al. *Curr. Topics in Micro. and Immunol.* (1992) 158:97-129). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (see for example Flotte et al. (1992) *Am. J. Respir. Cell. Mol. Biol.* 7:349-356; Samulski et al. (1989) *J. Virol.* 63:3822-3828; and McLaughlin et al. (1989) *J. Virol.* 62:1963-1973). Vectors containing as little as 300 base pairs of AAV can be packaged and can integrate. Space for exogenous DNA is limited to about 4.5 kb. An AAV vector such as that described in Tratschin et al. (1985) *Mol. Cell. Biol.* 5:3251-3260 can be used to introduce DNA into cells. A variety of nucleic acids have been introduced into different cell types using AAV vectors (see for example Hermonat et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6466-6470; Tratschin et al. (1985) *Mol. Cell. Biol.*

4:2072-2081; Wondisford et al. (1988) *Mol. Endocrinol.* 2:32-39; Tratschin et al. (1984) *J. Virol.* 51:611-619; and Flotte et al. (1993) *J. Biol. Chem.* 268:3781-3790).

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a PDF polypeptide in the tissue of an animal. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the subject *PDF* polypeptide gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In a representative embodiment, a gene encoding the subject *PDF* polypeptides can be entrapped in liposomes bearing positive charges on their surface (e.g., lipofectins) and (optionally) which are tagged with antibodies against cell surface antigens of the target tissue (Mizuno et al. (1992) *No Shinkei Geka* 20:547-551; PCT publication WO91/06309; Japanese patent application 1047381; and European patent publication EP-A-43075). For example, lipofection of cells can be carried out using liposomes tagged with monoclonal antibodies against any cell surface antigen present on the target cells.

In clinical settings, the gene delivery systems for the therapeutic *PDF* gene can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) *PNAS* 91: 3054-3057).

The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

Another aspect of the present invention concerns recombinant forms of the subject *PDF* protein which are encoded by genes derived from eukaryotic organisms such as mammals, e.g. humans. Recombinant proteins preferred by the present invention, in addition to native *PDF* polypeptides, are at least 70% homologous, more preferably at least 80% homologous and most preferably at least 85% homologous with an amino acid sequence shown in SEQ ID No: 2, 4 or 6. Polypeptides having an activity of the subject *PDF* polypeptides (i.e. either agonistic or antagonistic) and having at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology with a polypeptide sequence in SEQ ID No: 2, 4 or 6 are also within the scope of the invention.

The term "recombinant protein" refers to a polypeptide of the present invention which is produced by recombinant DNA techniques, wherein generally DNA encoding a PDF polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from", with respect to a recombinant *PDF* gene, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native *PDF* polypeptide, or an amino acid sequence similar thereto which is generated by mutations including substitutions and deletions (including truncation) of a naturally occurring form of a PDF protein. For instance, N-glycosylation sites in the *PDF* protein can be modified (e.g. mutated) to preclude glycosylation, allowing expression of a more homogenous, reduced carbohydrate analog in mammalian, insect and yeast expression systems. For example, the wild-type *PDF1* protein contains one potential N-linked glycosylation sites which can be mutated.

Furthermore, comparison of the *PDF1* with the other TGFβ superfamily receptor ligands known, namely BMP-2a and BMP-2b/BMP4 (Schlunegger, M.P. and Grutter, M.G. (1992) *Nature* 358:430-434), suggests that the biological activity of the molecule resides largely in the region of the protein containing the four cysteines whose spacing is apparently conserved so as to suggest an important motif, referred to herein as a "core motif". Consequently, a preferred *PDF1* polypeptide comprises a core motif bounded by the N and C-terminal cysteines of this four cysteine motif, e.g., of SEQ ID No. 2 or 6, or a sequence homologous thereto. Truncated forms of the subject *PDF* polypeptides, therefore, may be cropped at or in the vicinity of the C-terminus of the core motif.

The present invention further pertains to recombinant forms of the subject *PDF* polypeptides which are encoded by genes derived from a vertebrate organism, particularly a mammal (e.g. a human), and which have amino acid sequences evolutionarily related to the *PDF* proteins represented in SEQ ID No: 2, 4 or 6. Such

recombinant *PDF* polypeptides are preferably capable of functioning in one of either role of an agonist or antagonist of at least one biological activity of the *PDF* polypeptide of the appended sequence listing. The term "evolutionarily related to", with respect to amino acid sequences of the present recombinant *PDF* polypeptides, refers to *PDF* polypeptides having amino acid sequences which have arisen naturally, and also to mutational variants of *PDF* polypeptides which are derived, for example, by combinatorial mutagenesis. Such evolutionarily derived *PDF* polypeptides preferred by the present invention are at least 60% homologous, more preferably at least 70% homologous and most preferably at least 80% homologous with an amino acid sequence shown in SEQ ID No: 2, 4 or 6. Polypeptides having at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology with a sequence shown in SEQ ID No: 2, 4 or 6 are also within the scope of the invention.

The present invention further pertains to methods of producing the subject *PDF* polypeptides. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject *PDF* polypeptide can be cultured under appropriate conditions to allow expression of the peptide to occur. The peptide may be secreted and isolated from a mixture of cells and medium containing the recombinant *PDF* polypeptide. Alternatively, the peptide may be retained cytoplasmically by removing the signal peptide sequence from the recombinant *PDF* gene and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The recombinant *PDF* polypeptide peptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such peptide. In a preferred embodiment, the recombinant *PDF* polypeptide is a fusion protein containing a domain which facilitates its purification, such as a PDF/GST fusion protein.

This invention also pertains to a host cell transfected to express a recombinant form of the subject *PDF* polypeptides. The host cell may be any prokaryotic or eukaryotic cell, and the choice can be based at least in part on the desirablity of such post-translation modifications as glycosylation and/or addition of phosphatidylinositol. Thus, a nucleotide sequence derived from the cloning of *PDF*, encoding all or a selected portion of the full-length protein, can be used to produce a recombinant form of a PDF polypeptide via microbial or eukaryotic cellular processes. Ligating the polynucleotide sequence into a gene construct, such as an expression vector, and transforming or

transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known proteins, e.g. insulin, interferons, human growth hormone, IL-1, IL-2, and the like. Similar procedures, or modifications thereof, can be employed to prepare recombinant *PDF* polypeptides by microbial means or tissue-culture technology in accord with the subject invention.

The recombinant *PDF* gene can be produced by ligating nucleic acid encoding the subject *PDF* protein, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant forms of the subject *PDF* polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a PDF polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as *E. coli.*

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into *S. cerevisiae* (see, for example, Broach *et al.* (1983) in *Experimental Manipulation of Gene Expression*, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in *E. coli* due the presence of the pBR322 ori, and in *S. cerevisiae* due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used. In an illustrative embodiment, a PDF polypeptide is produced recombinantly utilizing an expression vector generated by sub-cloning the coding sequence of a PDF gene represented in SEQ ID NO. 2, 4 or 6.

The preferred mammalian expression vectors contain both prokaryotic sequences to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papillomavirus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression

systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17. In some instances, it may be desirable to express the recombinant *PDF* polypeptide by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the ß-gal containing pBlueBac III).

When it is desirable to express only a portion of a PDF protein, such as a form lacking a portion of the N-terminus, i.e. a truncation mutant which lacks the signal peptide, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from *E. coli* (Ben-Bassat et al. (1987) *J. Bacteriol.* 169:751-XX57) and *Salmonella typhimurium* and its *in vitro* activity has been demonstrated on recombinant proteins (Miller et al. (1987) *PNAS 84*:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either *in vivo* by expressing *PDF*-derived polypeptides in a host which produces MAP (e.g., *E. coli* or CM89 or *S. cerevisiae*), or *in vitro* by use of purified MAP (e.g., procedure of Miller et al., *supra*).

Alternatively, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. This type of expression system can be useful under conditions where it is desirable to produce an immunogenic fragment of a PDF protein. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of the *PDF* polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of a subject *PDF* protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising *PDF* epitopes as part of the virion. It has been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of a PDF protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP

Publication No: 0259149; and Evans et al. (1989) *Nature* 339:385; Huang et al. (1988) *J. Virol.* 62:3855; and Schlienger et al. (1992) *J. Virol.* 66:2).

The Multiple Antigen Peptide system for peptide-based immunization can also be utilized to generate an immunogen, wherein a desired portion of a PDF polypeptide is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett et al. (1988) *JBC* 263:1719 and Nardelli et al. (1992) *J. Immunol.* 148:914). Antigenic determinants of *PDF* proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression of proteins, including the *PDF* polypeptides of the present invention. For example, a PDF polypeptide can be generated as a glutathione-S-transferase (GST-fusion protein). Such GST-fusion proteins can enable easy purification of the *PDF* polypeptide, as for example by the use of glutathione-derivatized matrices (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. (N.Y.: John Wiley & Sons, 1991)). In another embodiment, a fusion gene coding for a purification leader sequence, such as a poly-(His)/enterokinase cleavage site sequence, can be used to replace the signal sequence which naturally occurs at N-terminus the *PDF* protein, in order to permit purification of the poly(His)-*PDF* protein by affinity chromatography using a Ni^{2+} metal resin. The purification leader sequence can then be subsequently removed by treatment with enterokinase (e.g., see Hochuli et al. (1987) *J. Chromatography* 411:177; and Janknecht et al. *PNAS* 88:8972).

Furthermore, the generation of *PDF* fusion proteins can be utilized as means for facilitating clustering, e.g., oligomerization, of *PDF* proteins to enhance certain activities associated with, for example, receptor cross-linking. The *PDF*/alkaline phosphatase fusion protein described herein may provide such a function, relying on the ability of alkaline phosphatase domains to promote complex formation between two or more *PDF*/AP proteins. Moreover, it may be desirable to provide multiple *PDF* domains in the same molecule, rather than rely on intermolecular complementation for oligomerization. For instance, an unstructured polypeptide linker region can be introduced between two *PDF* portions of the fusion protein. This linker can facilitate enhanced flexibility of the fusion protein, allowing the *PDF* domains to freely interact through intramolecular association, e.g., because of reduced steric hindrance between the two fragments, as well as permit appropriate folding of each fragment to occur. The linker can be of natural origin, such as a sequence determined to exist in random coil between two domains of a protein. Alternatively, the linker can be of synthetic origin. For instance, the sequence $(Gly_4Ser)_3$ can be used as a synthetic unstructured linker.

Linkers of this type are described in Huston et al. (1988) PNAS 85:4879; and U.S. Patent No. 5,091,513, both incorporated by reference herein.

Techniques for making fusion genes are known to those skilled in the art. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992).

The *PDF* polypeptide may also be chemically modified to create *PDF* derivatives by forming covalent or aggregrative conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of *PDF* can be prepared by linking the chemical moieties to functional groups on *PDF* amino acid sidechains or at the N-terminus or at the C-terminus of the polypeptide. For instance, a PDF protein can generated which includes a moiety, other than sequences naturally associated with the *PDF* protein, that binds a component of the extracellular matrix and enhances localization of the *PDF* analog to cell surfaces. For example, sequences derived from the fibronectin "type-III repeat", such as a tetrapeptide sequence R-G-D-S (Pierschbacher et al. (1984) *Nature* 309:30-3; and Kornblihtt et al. (1985) *EMBO* 4:1755-9) can be added to the *PDF* polypeptide to support attachment of the chimeric molecule to a cell through binding ECM components (Ruoslahti et al. (1987) *Science* 238:491-497; Pierschbacheret al. (1987) *J. Biol. Chem.* 262:17294-8.; Hynes (1987) *Cell* 48:549-54; and Hynes (1992) *Cell* 69:11-25) particularly where the *PDF* polypeptide lacks a C-terminal phosphatidylinositol.

The present invention also makes available isolated *PDF* polypeptides which are isolated from, or otherwise substantially free of other cellular and extracellular proteins, especially TGFβ or wnt family receptor proteins or other extracellular factors, normally associated with a *PDF* polypeptide. The term "substantially free of other cellular or extracellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of *PDF* polypeptides having less than 20% (by dry weight) contaminating protein, and

preferably having less than 5% contaminating protein. Functional forms of the subject *PDF* polypeptides can be prepared, for the first time, as purified preparations by using a cloned gene as described herein. Alternatively, the subject *PDF* polypeptides can be isolated by affinity purification using, for example, matrix bound TGFβ superfamily or wnt receptor protein. By "purified", it is meant, when referring to a peptide or DNA or RNA sequence, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins. The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (e.g., in an acrylamide gel) but not obtained either as pure (e.g. lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, e.g. acrylamide or agarose) substances or solutions.

As described above for recombinant polypeptides, isolated *PDF* polypeptides can include all or a portion of an amino acid sequence represented in SEQ ID No. 2, 4 or 6, or homologous sequences thereto. Exemplary derivatives of that sequence include proteins which lack N-glycosylation sites (e.g. to produce an unglycosylated protein), or which lack an N-terminus and or/C-terminus sequence, e.g. a PDF polypeptide which consists essentially of (with respect to receptor binding) a core sequence motif, or a sequence homologous thereto.

Furthermore, isolated peptidyl portions of *PDF* proteins can also be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. . In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a PDF polypeptide of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a PDF polypeptide activity, such as by *in vivo* competition assays or *in vitro* protein binding assays with *PDF* receptors.

It will also be possible to modify the structure of the subject *PDF* polypeptides for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, are considered functional equivalents of the *PDF* polypeptide described in more detail herein. Such modified peptide can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. conservative mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (see, for example, *Biochemistry*, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a peptide results in a functional *PDF* homolog (e.g. functional in the sense that it acts to mimic or antagonize the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type *PDF* protein or competitively inhibit such a response. Peptides in which more than one replacement has taken place can readily be tested in the same manner.

Accordingly, the present invention contemplates a method of generating sets of combinatorial mutants of the presently disclosed novel *PDF* polypeptides, as well as truncation and fragmentation mutants, and is especially useful for identifying potential variant sequences which are functional in binding to a PDF receptor. One purpose for screening such combinatorial libraries is, for example, to isolate novel *PDF* homologs which function as one of either an agonist or antagonist of the biological activities of the

wild-type ("authentic") protein, or alternatively, which possess novel activities all together. To illustrate, *PDF* homologs can be engineered by the present method to provide proteins which bind a TGFβ superfamily or wnt receptor, such as *PDF* receptors, yet which block receptor-mediated gene transcription resulting from signal transduction pathways normally associated with activation of that receptor. Such proteins, when expressed from recombinant DNA constructs, can be used in gene therapy protocols, or can be formulated as pharmaceutical preparations and delivered in such manner.

Likewise, mutagenesis can give rise to *PDF* homologs which have extracellular half-lives dramatically different than the corresponding wild-type protein. For example, the altered protein can be rendered either more stable or less stable to proteolytic degradation or other extracellular process which result in destruction of, or otherwise inactivation of, a PDF polypeptide. Such *PDF* homologs can be utilized to alter the envelope of bioavailabilty for a recombinant *PDF* protein by modulating, for example, the plasma half-life of the protein. For instance, a short half-life can give rise to more transient biological effects associated with a particular recombinant *PDF* polypeptide and can therefore allow tighter control of protein levels within or around a particular tissue. As above, such proteins, and particularly their recombinant nucleic acid constructs, can be used in gene therapy protocols as well as formulated into pharmaceutical preparations.

In an illustrative embodiment of this method, the amino acid sequences for a population of *PDF* homologs or other related proteins are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, *PDF* homologs from one or more species, e.g. murine and chicken, or *PDF* homologs from the same species but which differ due to mutation. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. In a preferred embodiment, the variegated library of *PDF* variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential *PDF* sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g. for phage display) containing the set of *PDF* sequences therein.

There are many ways by which the library of potential *PDF* homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the

desired set of potential *PDF* sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Alternatives to the above combinatorial mutagenesis also exist. For example, *PDF* homologs (both agonist and antagonist forms) can be generated and isolated from a library by screening using, for example, alanine scanning mutagenesis and the like (Ruf et al. (1994) *Biochemistry* 33:1565-1572; Wang et al. (1994) *J. Biol. Chem.* 269:3095-3099; Balint et al. (1993) *Gene* 137:109-118; Grodberg et al. (1993) *Eur. J. Biochem.* 218:597-601; Nagashima et al. (1993) *J. Biol. Chem.* 268:2888-2892; Lowman et al. (1991) *Biochemistry* 30:10832-10838; and Cunningham et al. (1989) *Science* 244:1081-1085), by linker scanning mutagenesis (Gustin et al. (1993) *Virology* 193:653-660; Brown et al. (1992) *Mol. Cell Biol.* 12:2644-2652; McKnight et al. (1982) *Science* 232:316); by saturation mutagenesis (Meyers et al. (1986) *Science* 232:613); by PCR mutagenesis (Leung et al. (1989) *Method Cell Mol Biol* 1:11-19); or by random mutagenesis (Miller et al. (1992) *A Short Course in Bacterial Genetics*, CSHL Press, Cold Spring Harbor, NY; and Greener et al. (1994) *Strategies in Mol Biol* 7:32-34).

A wide range of techniques are known in the art for screening gene products of combinatorial libraries, and for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of *PDF* homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the illustrative assays described below are amenable to high through-put analysis as necessary to screen large numbers of degenerate *PDF* sequences created by combinatorial mutagenesis techniques.

In one screening assay, the candidate *PDF* polypeptides are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind a PDF receptor protein via this gene product is detected in a "panning assay". For

instance, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled *PDF* receptor can be used to score for potentially functional *PDF* polypeptide homologs. For example, the AP-*PDF* fusion proteins described above, or the equivalent fluorescently labeled receptors, can be used to detect *PDF* homolog which retain receptor-binding activity. In the application of fluorescently labeled receptor, cells can be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, separated by a fluorescence-activated cell sorter.

In an alternate embodiment, the gene library is expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and fl are most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

In an illustrative embodiment, the recombinant phage antibody system (RPAS, Pharamacia Catalog number 27-9400-01) can be easily modified for use in expressing and screening *PDF* combinatorial libraries. For instance, the pCANTAB 5 phagemid of the RPAS kit contains the gene which encodes the phage gIII coat protein. The *PDF* combinatorial gene library, such as which may encode the Cys_4 motifs of the degenerate *PDF* library described above, can be cloned into the phagemid adjacent to the gIII signal sequence such that it will be expressed as a gIII fusion protein. After ligation, the phagemid is used to transform competent *E. coli* TG1 cells. Transformed cells are subsequently infected with M13KO7 helper phage to rescue the phagemid and its candidate *PDF* gene insert. The resulting recombinant phage contain phagemid DNA encoding a specific candidate *PDF*, and display one or more copies of the corresponding

fusion coat protein. The phage-displayed candidate *PDF* proteins which are capable of binding a PDF receptor are selected or enriched by panning. For instance, the phage library can be on glutathione-immobilized *PDF* receptor/GST fusion proteins to enrich for *PDF* homologs which retain an ability to bind a PDF receptor. The bound phage is then isolated, and if the recombinant phage express at least one copy of the wild type gIII coat protein, they will retain their ability to infect *E. coli*. Thus, successive rounds of reinfection of *E. coli*, and panning will greatly enrich for *PDF* homologs.

Each of these homologs can subsequently be screened for further biological activities in order to differentiate agonists and antagonists. For example, receptor-binding homologs isolated from the combinatorial library can be tested for their effect on cellular proliferation relative to the wild-type form of the protein. Alternatively, one could screen the homologs for agonists by detecting autophosphorylation of a PDF receptor in response to treatment with the homolog (see, for example, Millauer et al. (1993) *Cell* 72:835-846). In similar fashion, antagonists can be identified from the enriched fraction based on their ability to inhibit autophosphorylation of the receptor by wild-type *PDF* protein.

In another embodiment, the combinatorial library is designed to be extracellularly presented (e.g. as it occurs naturally) and, though optionally, secreted (e.g. the polypeptides of the library all include a signal sequence). The gene can be used to transfect a eukaryotic cell that can be co-cultured with cells which express an functional TBFβ superfamily receptor, e.g. a PDF receptor, and which is sensitive to treatment with the wild-type soluble form of *PDF*. Functional *PDF* homologs secreted by the cells expressing the combinatorial library will diffuse to neighboring *PDF* receptor positive cells and induce a phenotypic change. Using, for example, antibodies directed to epitopes which are either created or destroyed in response to *PDF* treatment, the pattern of detection of *PDF* induction will resemble a gradient function, and will allow the isolation (generally after several repetitive rounds of selection) of cells producing active *PDF* homologs. Likewise, *PDF* antagonists can be selected in similar fashion by the ability of the cell producing a functional antagonist to protect neighboring cells from the effect of authentic *PDF* added to the culture media.

To illustrate, target cells are cultured in 24-well microtitre plates. The target cells can be, for instance, cells which naturally express *PDF*-related receptors, such as NIH 3T3 cells, or cells which have been transfected with genes encoding such a receptor. COS cells are transfected with the combinatorial *PDF* gene library and cultured (optionally) in a cell culture insert (e.g. Collaborative Biomedical Products, Catalog #40446) that are able to fit into the wells of the microtitre plate. The cell culture inserts

are placed in the wells such that recombinant *PDF* homologs secreted by the cells in the insert can diffuse through the porous bottom of the insert and contact the target cells in the microtitre plate wells. After a period of time sufficient for functional forms of *PDF* to produce a measurable response in the target cells, the inserts are removed and the effect of any *PDF* homologs on the target cells determined. Cells from the inserts corresponding to wells which score positive for activity can be split and re-cultured on several inserts, the process being repeated until the active clones are identified.

The invention also provides for reduction of the *PDF* polypeptides to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt binding of a PDF polypeptide of the present invention with a PDF receptor. Accordingly, such mutagenic techniques as described above are also useful to map the determinants of the *PDF* polypeptides which participate in protein-protein interactions involved in, for example, binding of the subject *PDF* polypeptide to a PDF receptor or in causing oligomerization of receptors. To illustrate, the critical residues of a subject *PDF* polypeptide which are involved in molecular recognition of a PDF receptor can be determined and used to generate *PDF* polypeptide-derived peptidomimetics which competitively inhibit binding of the authentic *PDF* protein with that receptor. By employing, for example, scanning mutagenesis to map the amino acid residues of the *PDF* protein involved in binding the *PDF* receptor, (for example the amino acids preferred fragments comprise the nucleic acids which encode the amino acid residues CKKTPLTIDFKEIG) peptidomimetic compounds can be generated which mimic those residues in binding to the receptor and which consequently can inhibit binding of *PDF* to its cognate receptor and interfere with the function of the receptor. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gama lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and β-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun*126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

Another aspect of the invention pertains to an antibody specifically reactive with a PDF protein. For example, by using immunogens derived from a PDF protein, e.g. based on the cDNA sequences, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal, such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., a PDF polypeptide or an antigenic fragment which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the *PDF* protein can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the *PDF* protein of the present invention, e.g. antigenic determinants of a protein represented by SEQ ID No: 2, 4 or 6 or a closely related human or non-human mammalian homolog (e.g. at least 85 percent homologous, preferably at least 90 percent homologous, and more preferably at least 95 percent homologous). In yet a further preferred embodiment of the present invention, the anti-*PDF* polypeptide antibodies do not substantially cross react (i.e. does not react specifically) with a protein which is, for example, less than 85 percent homologous to SEQ ID No: 2, 4 or 6; or less than 95 percent homologous with SEQ ID No: 2, 4 or 6; or less than 98-99 percent homologous with SEQ ID No: 2, 4 or 6. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein (e.g. other proteins of the TGF-β superfamily) which is at least one order of magnitude, more preferably at least 2 orders of magnitude, and even more preferably at least 3 orders of magnitude less than the binding affinity of the antibody for the protein of SEQ ID No: 2, 4 or 6.

Following immunization, anti-*PDF* antisera can be obtained and, if desired, polyclonal anti-*PDF* antibodies isolated from the serum. To produce monoclonal antibodies, antibody-producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, an include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) *Nature*, 256: 495-497), the human B cell hybridoma technique (Kozbar et al., (1983) *Immunology Today*, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened

immunochemically for production of antibodies specifically reactive with a PDF polypeptide of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with one of the subject *PDF* polypeptide. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab)_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab)_2$ fragment can be treated to reduce disulfide bridges to produce Fab fragments. The antibody of the present invention is further intended to include single chain, bispecific and chimeric molecules having a PDF affinity conferred by at least one CDR region of the antibody.

Both monoclonal and polyclonal antibodies (Ab) directed against *PDF* polypeptide or *PDF* polypeptide variants, and antibody fragments such as Fab and $F(ab)_2$, can be used to block the action of *PDF* and allow the study of the role of *PDF* in, for example, embryogenesis and/or tumorogenesis. For example, purified monoclonal Abs can be injected directly into the limb buds of chick or mouse embryos. It is demonstrated in the examples below that *PDF* is expressed in the limb buds of day 10.5 embryos. Thus, the use of anti-*PDF* Abs during this developmental stage can allow assessment of the effect of *PDF* on the formation of limbs *in vivo*. In a similar approach, hybridomas producing anti-*PDF* monoclonal Abs, or biodegradable gels in which anti-*PDF* Abs are suspended, can be implanted at a site proximal or within the area at which *PDF* action is intended to be blocked. Experiments of this nature can aid in deciphering the role of this and other factors that may be involved in limb patterning and tissue formation.

Antibodies which specifically bind *PDF* polypeptide epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of each of the subject *PDF* polypeptides. Anti-*PDF* antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *PDF* protein levels in tissue or bodily fluid as part of a clinical testing procedure. For instance, such measurements can be useful in predictive valuations of the onset or progression of disorders, involving tissue derived from embryonic endoderm, such as diabetes. Likewise, the ability to monitor *PDF* levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of *PDF* polypeptides can be measured in bodily fluid, such as in samples of cerebral spinal fluid, or can be measured in tissue, such as produced by biopsy. Diagnostic assays using anti-*PDF* antibodies can include, for

example, immunoassays designed to aid in early diagnosis of a neurodegenerative disorder, particularly ones which are manifest at birth. Diagnostic assays using anti-*PDF* polypeptide antibodies can also include immunoassays designed to aid in early diagnosis and phenotyping of a neoplastic or hyperplastic disorder.

Another application of anti-*PDF* antibodies of the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λgt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of β-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a PDF protein can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*PDF* antibodies. Positive phage detected by this assay can then be isolated from the infected plate. Thus, the presence of *PDF* homologs (orthologs) can be detected and cloned from other animals, as can alternate isoforms (including splicing variants).

Moreover, the nucleotide sequence determined from the cloning of the *PDF* gene will further allow for the generation of probes and primers designed for use in identifying and/or cloning *PDF* homologs in other cell types, e.g. from other tissues, as well as *PDF* homologs from other animals, e.g. humans. For instance, the present invention also provides a probe/primer comprising a substantially purified oligonucleotide, which oligonucleotide comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or anti-sense sequence of SEQ ID No: 1, 3 or 5 or naturally occurring mutants thereof. For instance, primers based on the nucleic acids represented in SEQ ID No. 1, 3 or 5 can be used in PCR reactions to clone *PDF* homologs. Likewise, probes based on the *PDF* gene sequence of SEQ ID No. 1, 3 or 5 can be used to detect *PDF* transcripts or genomic sequences. In preferred embodiments, the probe further comprises a label group attached thereto and able to be detected, e.g. the label group is selected from the group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors. Such probes can also be used as a part of a diagnostic test kit for identifying cells in which *PDF* is misexpressed, such as by measuring a level of a PDF encoding nucleic acid in a sample of cells from a patient; e.g. detecting *PDF* mRNA levels or determining whether a genomic *PDF* gene has been mutated or deleted.

To illustrate, nucleotide probes can be generated from the *PDF* gene which facilitate histological screening of intact tissue and tissue samples for the presence of a PDF polypeptide mRNA. Similar to the diagnostic uses of anti-*PDF* polypeptide

antibodies, the use of probes directed to *PDF* messages, or to genomic *PDF* sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. unwanted cell growth) or abnormal differentiation of tissue. Used in conjunction with anti-*PDF* immunoassays, the nucleotide probes can help facilitate the determination of the molecular basis for a developmental disorder which may involve some abnormality associated with expression (or lack thereof) of a PDF polypeptide. For instance, variation in *PDF* polypeptide synthesis can be differentiated from a mutation in the *PDF* coding sequence.

Accordingly, the present method provides a method for determining if a subject is at risk for a disorder characterized by unwanted cell proliferation or abherent control of differentiation. In preferred embodiments, the subject method can be generally characterized as comprising detecting, in a tissue sample of the subject (e.g. a human patient), the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a PDF polypeptide or (ii) the mis-expression of a PDF gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a PDF gene, (ii) an addition of one or more nucleotides to such a PDF gene, (iii) a substitution of one or more nucleotides of a PDF gene, (iv) a gross chromosomal rearrangement of a PDF genes, (v) a gross alteration in the level of a messenger RNA transcript of a PDF gene, (vi) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a PDF gene, and (vii) a non-wild type level of a PDF polypeptide. In one aspect of the invention there is provided a probe/primer comprising an oligonucleotide containing a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of SEQ ID No: 1, 3 or 5, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with a PDF gene. The probe is exposed to nucleic acid of a tissue sample; and the hybridization of the probe to the sample nucleic acid is detected. In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent No: 4,683,195 and 4,683,202) or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science*, 241:1077-1080; and NaKazawa et al. (1944) *PNAS* 91:360-364) the later of which can be particularly useful for detecting point mutations in the *PDF* gene. Alternatively, immunoassays can be employed to determine the level of *PDF* protein, either soluble or membrane bound.

Also, the use of anti-sense techniques (e.g. microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to a PDF

mRNA or gene sequence) can be used to investigate role of *PDF* in developmental events, as well as the normal cellular function of *PDF* in adult tissue. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals.

Furthermore, by making available purified and recombinant *PDF* polypeptides, the present invention facilitates the development of assays which can be used to screen for drugs, or for *PDF* homologs, which are either agonists or antagonists of the normal cellular function of the subject *PDF* polypeptides, or of their role in the pathogenesis of cellular proliferation and/or differentiation and disorders related thereto. In one embodiment, the assay evaluates the ability of a compound to modulate binding between a PDF polypeptide and a PDF receptor. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by skilled artisan.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with receptor proteins. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with a PDF receptor polypeptide which is ordinarily capable of binding a PDF protein. To the mixture of the compound and receptor is then added a composition containing a *PDF* polypeptide. Detection and quantification of receptor/*PDF* complexes provides a means for determining the compound's efficacy at inhibiting (or potentiating) complex formation between the receptor protein and the *PDF* polypeptide. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified *PDF* polypeptide is added to a composition containing the receptor protein, and the formation of receptor/*PDF* complex is quantitated in the absence of the test compound.

Complex formation between the *PDF* polypeptide and a PDF receptor may be detected by a variety of techniques. For instance, modulation of the formation of complexes can be quantitated using, for example, detectably labelled proteins such as

radiolabelled, fluorescently labelled, or enzymatically labelled *PDF* polypeptides, by immunoassay, or by chromatographic detection.

Typically, it will be desirable to immobilize either the *PDF* receptor or the *PDF* polypeptide to facilitate separation of receptor/*PDF* complexes from uncomplexed forms of one of the proteins, as well as to accomadate automation of the assay. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/receptor (GST/receptor) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the *PDF* polypeptide, e.g. an ^{35}S-labeled *PDF* polypeptide, and the test compound and incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired, e.g., at 4°C in a buffer containing 0.6M NaCl or a detergent such as 0.1% Triton X-100. Following incubation, the beads are washed to remove any unbound *PDF* polypeptide, and the matrix bead-bound radiolabel determined directly (e.g. beads placed in scintilant), or in the supernatant after the receptor/*PDF* complexes are dissociated. Alternatively, the complexes can dissociated from the bead, separated by SDS-PAGE gel, and the level of *PDF* polypeptide found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, the *PDF* receptor protein can be immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated receptor molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the *PDF* receptor but which do not interfere with *PDF* binding can be derivatized to the wells of the plate, and the receptor trapped in the wells by antibody conjugation. As above, preparations of a PDF polypeptide and a test compound are incubated in the receptor-presenting wells of the plate, and the amount of receptor/*PDF* complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the *PDF* polypeptide, or which are reactive with the receptor protein and compete for binding with the *PDF* polypeptide; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the *PDF* polypeptide. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with the *PDF*

polypeptide. To illustrate, the *PDF* polypeptide can be chemically cross-linked or genetically fused with alkaline phosphatase, and the amount of *PDF* polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. paranitrophenylphosphate. Likewise, a fusion protein comprising the *PDF* polypeptide and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) *J Biol Chem* 249:7130).

For processes which rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein, such as the anti-*PDF* antibodies described herein, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the *PDF* polypeptide or *PDF* receptor sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) *J Biol Chem* 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharamacia, NJ).

Another aspect of the present invention relates to a method of inducing and/or maintaining a differentiated state, enhancing survival, and/or promoting proliferation of a cell responsive to a PDF protein, by contacting the cells with a PDF agonist or a PDF antagonist. For instance, it is contemplated by the invention that, in light of the present finding of an apparently broad involvement of *PDF* proteins in the formation of ordered spatial arrangements of differentiated tissues in vertebrates, the subject method could be used to generate and/or maintain an array of different vertebrate tissue both *in vitro* and *in vivo*. The *PDF* agent can be, as appropriate, any of the preparations described above, including isolated polypeptides, gene therapy constructs, antisense molecules, peptidomimetics or agents identified in the drug assays provided herein.

Another aspect of the present invention relates to a method of inducing and/or maintaining a differentiated state, enhancing survival, and/or promoting (or alternatively inhibiting) proliferation of a cell responsive to a TGF-β factor, by contacting the cells with an agent which modulates *PDF*-dependent signaling. For instance, it is contemplated by the invention that, in light of the present finding of an apparent involvement of *PDF* proteins in the formation of ordered spatial arrangements of differentiated tissues in mammals, the subject method could be used to generate and/or maintain an array of different tissue both *in vitro* and *in vivo*. A "*PDF* therapeutic,"

whether inductive or anti-inductive with respect to signaling by a TGF-β or wnt protein, can be, as appropriate, any of the preparations described above, including isolated polypeptides, gene therapy constructs, antisense molecules, peptidomimetics or agents identified in the drug assays provided herein. Moreover, it is contemplated that, based on the observation of activity of the vertebrate for purposes of therapeutic and diagnostic uses, may include invertebrate homologs of *PDF*.

There are a wide variety of pathological cell proliferative conditions for which *PDF* therapeutics of the present invention can be used in treatment. For instance, such agents can provide therapeutic benefits where the general strategy being the inhibition of an anomalous cell proliferation. Diseases that might benefit from this methodology include, but are not limited to diabetes, liver disease, various cancers and leukemias, psoriasis, bone diseases, fibroproliferative disorders such as involving connective tissues, atherosclerosis and other smooth muscle proliferative disorders, as well as chronic inflammation.

In addition to proliferative disorders, the present invention contemplates the use of *PDF* therapeutics for the treatment of differentiative disorders which result from, for example, de-differentiation of tissue which may (optionally) be accompanied by abortive reentry into mitosis, e.g. apoptosis. Such degenerative disorders include chronic neurodegenerative diseases of the nervous system, including Alzheimer's disease, Parkinson's disease, Huntington's chorea, amylotrophic lateral sclerosis and the like, as well as spinocerebellar degenerations. Other differentiative disorders include, for example, disorders associated with connective tissue, such as may occur due to de-differentiation of chondrocytes or osteocytes, as well as vascular disorders which involve de-differentiation of endothelial tissue and smooth muscle cells, gastric ulcers characterized by degenerative changes in glandular cells, and renal conditions marked by failure to differentiate, e.g. Wilm's tumors.

It will also be apparent that, by transient use of modulators of *PDF* pathways, *in vivo* reformation of tissue can be accomplished, e.g. in the development and maintenance of organs. By controlling the proliferative and differentiative potential for different cells, the subject gene constructs can be used to reform injured tissue, or to improve grafting and morphology of transplanted tissue. For instance, *PDF* agonists and antagonists can be employed in a differential manner to regulate different stages of organ repair after physical, chemical or pathological insult. For example, such regimens can be utilized in repair of pancreatic endocrine or exocrine function, cartilage repair, increasing bone density, liver repair subsequent to a partial hepatectomy, or to promote regeneration of lung tissue in the treatment of emphysema.

In one embodiment, the present invention makes use of the discovery that *PDF* proteins are likely to be involved in controlling the development and formation of the digestive tract, liver, pancreas, lungs, kidney, and other organs which derive from the primitive gut. *PDF2* has also been shown to play a role in the development of the eye. As described in the Examples below, *PDF* proteins are presumptively involved in cellular activity as TGF-β–like or wnt-like inductive signals. Accordingly, *PDF* agonists and/or antagonists can be employed in the development and maintenance of an artificial liver which can have multiple metabolic functions of a normal liver. In an exemplary embodiment, the *PDF* therapeutic can be used in a method for treating a disorder characterized by insufficient liver function. For example, stem cells can be induced to differentiate into liver cells. The *PDF* therapeutic can be used alone, or can be used in combination with other factors which act to more particularly enhance a particular differentiation fate of the hepatic progenitor cell. In the later instance, a *PDF* therapeutic might be viewed as ensuring that the treated cell has achieved a particular phenotypic state such that the cell is poised along a certain developmental pathway so as to be properly induced upon contact with a secondary growth factor. In an exemplary embodiment, *PDF* therapeutics can be used to induce and/or maintain differentiation of digestive tube stem cells to form hepatocyte cultures which can be used to populate extracellular matrices, or which can be encapsulated in biocompatible polymers, to form both implantable and extracorporeal artificial livers.

In another embodiment, compositions of *PDF* therapeutics can be utilized in conjunction with transplantation of such artificial livers, as well as embryonic liver structures, to promote intraperitoneal implantation, vascularization, and *in vivo* differentiation and maintenance of the engrafted liver tissue.

Similar utilization of *PDF* therapeutics are contemplated in the generation and maintenance of pancreatic cultures and artificial pancreatic tissues and organs. In yet another preferred embodiment, the pancreatic progenitor cells are induced to differentiate into pancreatic islet cells, e.g., β islet cells, α islet cells, δ islet cells, or ϕ islet cells, subsequent to being introduced into the subject. Preferably, the pancreatic progenitors cells are induced to differentiate into pancreatic islet, e.g., β islet cells, α islet cells, δ islet cells, or ϕ islet cells, in culture prior to introduction into the subject.

There are a wide variety of pathological cell proliferative and differentiative conditions for which the *PDF* gene constructs of the present invention may provide therapeutic benefits, with the general strategy being, for example, the correction of abherent insulin expression, or modulation of differentiative events mediated by *PDF*, such as may be influenced by transcriptional regulatory sequences of other genes with

which the subject *PDF* interact. More generally, however, the present invention relates to a method of inducing and/or maintaining a differentiated state, enhancing survival and/or affecting proliferation of a cell in which *PDF* responsive genes are expressed, by contacting the cell with an agent which modulates the function (as an agonist or an antagonist) of *PDF*. For instance, it is contemplated by the invention that, in light of the apparent involvement of *PDF* in the formation of ordered spatial arrangements of pancreatic tissues, the subject method could be used to generate and/or maintain such tissue both *in vitro* and *in vivo*. For instance, modulation of the function of *PDF* can be employed in both cell culture and therapeutic methods involving generation and maintenance β-cells and possibly also for non-pancreatic tissue, such as in controlling the development and maintenance of tissue from the digestive tract, spleen, lungs, and other organs which derive from the primitive gut. The agent can be, as appropriate, any of the preparations described herein, including gene therapy constructs, antisense molecules, peptidomimetics or other agents identified in the drug screening assays provided herein.

In yet another exemplary embodiment, the present method can be used in the treatment of hyperplastic and neoplastic disorders effecting pancreatic tissue, particularly those characterized by abherent proliferation of β-cells, or mis-expression of *PDF* or other proteins involved in regulatory complexes involving *PDF*. For instance, pancreatic cancers are marked by abnormal proliferation of pancreatic cells which can result in alterations of insulin secretory capacity of the pancreas. For instance, certain pancreatic hyperplasias, such as pancreatic carcinomas, can result in hypoinsulinemia due to dysfunction of β-cells or decreased islet cell mass. Stimulation of *PDF*-mediated expression of insulin, such as by overexpression of exogenous *PDF* in β-cells, can be used to increase the insulin production of normal β-cells in the tissue, as well as enhance regeneration of the tissue after anti-tumor therapy.

In contrast, other pancreatic tumors, such as islet tumors (e.g., insulinomas), are marked by overproduction of insulin (i.e., hyperinsulinemia), which can cause hypoglycemic conditions in a patient. Indeed, hypoglycemia can result from any one of a number of different disorders which result in raised plasma insulin levels, including other β-cell abnormalities, as well as endocrinopathies, sepsis (including malaria), congestive cardiac failure, hepatic and renal insufficiencies, various genetic abnormalities of metabolism, and exogenous toxins (such as alcohol). According to the present invention, hypoglycemic conditions can be treated by administering therapeutic amounts of an agent able to antagonize *PDF*-mediated expression of insulin. Depending on the desired half-life of the effects of the treatment, such agents can range from

peptidomimetic and other small molecule inhibitors of *PDF* function, to antisense constructs, to transient or long-term gene therapy regimens.

Furthermore, the subject method can be used as part of treatments for various forms of diabetes, as well as other pathologies resulting from direct physical/chemical damage to β-cells which result in necrosis and loss of functional islet tissue. In diabetes mellitus, insulin secretion is either completely absent (IDDM) or inappropriately regulated (NIDDM). However, each is characterized by the presence of chronically elevated levels of blood glucose (hyperglycemia). The primary aim of treatment in both forms is the same, namely, the reduction of blood glucose levels to as near as normal as possible. For example, treatment of IDDM typically involves administration of replacement doses of insulin. In contrast, initial therapy for NIDDM may be based in part on therapies which include administration of hypoglycemic agents such as sulfonylurea, though insulin treatment in later stages of the disease may be required to effect normoglycemia. Accordingly, the present method can provide a means for controlling diabetogenous glycemic levels, by administeration of a PDF agonist (e.g. a hyperglycemic agent) as, for example, by causing recombinant expression of a wild-type form of the protein in β-islet cells of the patient, or alternatively, administration of a PDF antagonist (e.g. a hypoglycemic agent) such as a molecule which inhibits response element binding and/or activation of insulin gene transcription by *PDF* or *PDF*-containing complexes.

Moreover, manipulation of *PDF*-mediated gene expression, such as of the insulin gene, may be useful for reshaping/repairing pancreatic tissue both *in vivo* and *in vitro*. In one embodiment, the present invention makes use of the apparent involvement of the subject *PDF* protein in regulating the development of pancreatic tissue responsible for formation of β-cells, e.g. induction of β-cell differentiation from ductal tissue, as well as other tissue from the lungs and other organs which derive from the primitive gut. For example, therapeutic compositions for modulating the role of *PDF* in tissue differentiation can be utilized to preserve any β-cells that have not been destroyed by diabetic or tumorogenic causes, as well as to induce regeneration of β-cells so as to increase the islet mass. In general, the subject method can be employed therapeutically to regulate the pancreas after physical, chemical or pathological insult.

In yet another embodiment, the subject method can be applied to cell culture techniques, and in particular, may be employed to enhance the initial generation of prosthetic pancreatic tissue devices. Manipulation of *PDF* function, for example, by altering the ability of the protein to transactivate *PDF* responsive genes, can provide a means for more carefully controlling the characteristics of a cultured tissue. In an

exemplary embodiment, the subject method can be used to augment production of prosthetic devices which require β-islet cells, such as may be used in the encapsulation devices described in, for example, the Aebischer et al. U.S. Patent No. 4,892,538, the Aebischer et al. U.S. Patent No. 5,106,627, the Lim U.S. Patent No. 4,391,909, and the Sefton U.S. Patent No. 4,353,888. Early progenitor cells to the pancreatic islets are multipotential, and apparently coactivates all the islet-specific genes from the time they first appear. As development proceeds, expression of islet-specific hormones, such as insulin, becomes restricted to the pattern of expression characteristic of mature islet cells. The phenotype of mature islet cells, however, is not stable in culture, as reappearance of embryonal traits in mature β-cells can be observed.

Furthermore, manipulation of the differentiative state of pancreatic tissue can be utilized in conjunction with transplantation of artificial pancreas so as to promote implantation, vascularization, and *in vivo* differentiation and maintenance of the engrafted tissue. For instance, manipulation of *PDF* function to affect tissue differentiation can be utilized as a means of maintaining graft viability.

In another embodiment, *in vitro* cell cultures can be used for the identification, isolation, and study of genes and gene products that are expressed in response to disruption of *PDF*-mediated signal transduction, and therefore likely involved in development and/or maintenance of tissues. These genes would be "downstream" of the *PDF* gene product. For example, if new transcription is required for *PDF*-mediated induction, a subtractive cDNA library prepared with control cells and cells overexpressing a *PDF* gene can be used to isolate genes that are turned on or turned off by this process. The powerful subtractive library methodology incorporating PCR technology described by Wang and Brown is an example of a methodology useful in conjunction with the present invention to isolate such genes (Wang et al. (1991) Proc. Natl. Acad. Sci. USA 88:11505-11509). For example, this approach has been used successfully to isolate more than sixteen genes involved in tail resorption with and without thyroid hormone treatment in Xenopus. Utilizing control and treated cells, the induced pool can be subtracted from the uninduced pool to isolate genes that are turned on, and then the uninduced pool from the induced pool for genes that are turned off. From this screen, it is expected that two classes of mRNAs can be identified. Class I RNAs would include those RNAs expressed in untreated cells and reduced or eliminated in induced cells, that is the down-regulated population of RNAs. Class II RNAs include RNAs that are upregulated in response to induction and thus more abundant in treated than in untreated cells. RNA extracted from treated vs. untreated cells can be used as a primary test for the classification of the clones isolated from the libraries. Clones of each

class can be further characterized by sequencing and, their spatiotemporal distribution determined in the embryo by whole mount *in situ* and developmental northern blots analysis.

In yet another embodiment, *PDF* therapeutics can be employed to regulate such organs after physical, chemical or pathological insult. For instance, therapeutic compositions comprising *PDF* therapeutics can be utilized in liver repair subsequent to a partial hepatectomy. Similarly, therapeutic compositions containing *PDF* therapeutics can be used to promote regeneration of lung tissue in the treatment of emphysema.

In still another embodiment of the present invention, compositions comprising *PDF* therapeutics can be used for the *in vitro* generation of skeletal tissue, such as from skeletogenic stem cells, as well as for the *in vivo* treatment of skeletal tissue deficiencies. The present invention particularly contemplates the use of *PDF* therapeutics which upregulate or mimic the inductive activity of a bone morphogenetic protein (BMP) or TGF-β, such as may be useful to control chondrogenesis and/or osteogenesis. By "skeletal tissue deficiency", it is meant a deficiency in bone or other skeletal connective tissue at any site where it is desired to restore the bone or connective tissue, no matter how the deficiency originated, e.g. whether as a result of surgical intervention, removal of tumor, ulceration, implant, fracture, or other traumatic or degenerative conditions, so long as modulation of a TGF-β inductive response is appropriate.

For instance, the present invention makes available effective therapeutic methods and *PDF* therapeutic compositions for restoring cartilage function to a connective tissue. Such methods are useful in, for example, the repair of defects or lesions in cartilage tissue which is the result of degenerative wear such as that which results in arthritis, as well as other mechanical derangements which may be caused by trauma to the tissue, such as a displacement of torn meniscus tissue, meniscectomy, a laxation of a joint by a torn ligament, malignment of joints, bone fracture, or by hereditary disease. The present reparative method is also useful for remodeling cartilage matrix, such as in plastic or reconstructive surgery, as well as periodontal surgery. The present method may also be applied to improving a previous reparative procedure, for example, following surgical repair of a meniscus, ligament, or cartilage. Furthermore, it may prevent the onset or exacerbation of degenerative disease if applied early enough after trauma.

In one embodiment of the present invention, the subject method comprises treating the afflicted connective tissue with a therapeutically sufficient amount of a *PDF* therapeutic to generate a cartilage repair response in the connective tissue by stimulating the differentiation and/or proliferation of chondrocytes embedded in the tissue.

Induction of chondrocytes by treatment with a *PDF* therapeutic can subsequently result in the synthesis of new cartilage matrix by the treated cells. Such connective tissues as articular cartilage, interarticular cartilage (menisci), costal cartilage (connecting the true ribs and the sternum), ligaments, and tendons are particularly amenable to treatment in reconstructive and/or regenerative therapies using the subject method. As used herein, regenerative therapies include treatment of degenerative states which have progressed to the point of which impairment of the tissue is obviously manifest, as well as preventive treatments of tissue where degeneration is in its earliest stages or imminent. The subject method can further be used to prevent the spread of mineralization into fibrotic tissue by maintaining a constant production of new cartilage.

In an illustrative embodiment, the subject method can be used to treat cartilage of a diarthroidal joint, such as a knee, an ankle, an elbow, a hip, a wrist, a knuckle of either a finger or toe, or a temperomandibular joint. The treatment can be directed to the meniscus of the joint, to the articular cartilage of the joint, or both. To further illustrate, the subject method can be used to treat a degenerative disorder of a knee, such as which might be the result of traumatic injury (e.g., a sports injury or excessive wear) or osteoarthritis. An injection of a *PDF* therapeutic into the joint with, for instance, an arthroscopic needle, can be used to treat the afflicted cartilage. In some instances, the injected agent can be in the form of a hydrogel or other slow release vehicle described above in order to permit a more extended and regular contact of the agent with the treated tissue.

In one embodiment of the subject method, the implants are contacted with a *PDF* therapeutic during the culturing process so as to induce and/or maintain differentiated chondrocytes in the culture in order to further stimulate cartilage matrix production within the implant. In such a manner, the cultured cells can be caused to maintain a phenotype typical of a chondrogenic cell (i.e. hypertrophic), and hence continue the population of the matrix and production of cartilage tissue.

In another embodiment, the implanted device is treated with a *PDF* therapeutic in order to actively remodel the implanted matrix and to make it more suitable for its intended function. As set out above with respect to tissue transplants, the artificial transplants suffer from the same deficiency of not being derived in a setting which is comparable to the actual mechanical environment in which the matrix is implanted. The activation of the chondrocytes in the matrix by the subject method can allow the implant to acquire characteristics similar to the tissue for which it is intended to replace.

In yet another embodiment, the subject method is used to enhance attachment of prosthetic devices. To illustrate, the subject method can be used in the implantation of a periodontal prosthesis, wherein the treatment of the surrounding connective tissue stimulates formation of periodontal ligament about the prosthesis, as well as inhibits formation of fibrotic tissue proximate the prosthetic device.

In still further embodiments, the subject method can be employed for the generation of bone (osteogenesis) at a site in the animal where such skeletal tissue is deficient. TGF-β's, especially BMPs, are particularly associated with the hypertrophic chondrocytes that are ultimately replaced by osteoblasts as well as the production of bone matrix by osteocytes. Consequently, administration of a *PDF* therapeutic can be employed as part of a method for treating bone loss in a subject, e.g. to prevent and/or reverse osteoporosis and other osteopenic disorders, as well as to regulate bone growth and maturation. For example, preparations comprising *PDF* agonists can be employed, for example, to induce endochondral ossification by mimicking or potentiating the activity of a BMP, at least so far as to facilitate the formation of cartilaginous tissue precursors to form the "model" for ossification. Therapeutic compositions of *PDF* agonists can be supplemented, if required, with other osteoinductive factors, such as bone growth factors (e.g. TGF-β factors, such as the bone morphogenetic factors *BMP-2* and *BMP-4*, as well as activin), and may also include, or be administered in combination with, an inhibitor of bone resorption such as estrogen, bisphosphonate, sodium fluoride, calcitonin, or tamoxifen, or related compounds.

For certain cell-types, particularly in epithelial and hemopoietic cells, normal cell proliferation is marked by responsiveness to negative autocrine or paracrine growth regulators, such as members of the TGFβ family. This is generally accompanied by differentiation of the cell to a post-mitotic phenotype. However, it has been observed that a significant percentage of human cancers derived from these cells types display a reduced responsiveness to growth regulators such as TGFβ. For instance, some tumors of colorectal, liver epithelial, and epidermal origin show reduced sensitivity and resistance to the growth-inhibitory effects of TGFβ as compared to their normal counterparts. In this context, a noteworthy characteristic of several such transformed cell lines is the absence of detectable TGFβ receptors. Treatment of such tumors with *PDF* therapeutics provides an opportunity to mimic the effective function of TGFβ-mediated inhibition.

To further illustrate the use of the subject method, the therapeutic application of a *PDF* therapeutic can be used in the treatment of a neuroglioma. Gliomas account for 40-50% of intracranial tumors at all ages of life. Despite the increasing use of radiotherapy,

chemotherapy, and sometimes immunotherapy after surgery for malignant glioma, the mortality and morbidity rates have not substantially improved. However, there is increasing experimental and clinical evidence that for a significant number of gliomas, loss of TGFβ responsiveness is an important event in the loss of growth control. Where the cause of decreased responsiveness is due to loss of receptor or loss of other TGFβ signal transduction proteins upstream of a *PDF*, treatment with a *PDF* therapeutic can be used effectively to inhibit cell proliferation.

The subject *PDF* therapeutics can also be used in the treatment of hyperproliferative vascular disorders, e.g. smooth muscle hyperplasia (such as atherosclerosis) or restinosis, as well as other disorders characterized by fibrosis, e.g. rheumatoid arthritis, insulin dependent diabetes mellitus, glomerulonephritis, cirrhosis, and scleroderma, particularly proliferative disorders in which loss of a TGFβ autocrine or paracrine signaling is implicated.

For example, restinosis continues to limit the efficacy of coronary angioplasty despite various mechanical and pharmaceutical interventions that have been employed. An important mechanism involved in normal control of intimal proliferation of smooth muscle cells appears to be the induction of autocrine and paracrine TGFβ inhibitory loops in the smooth muscle cells (Scott-Burden et al. (1994) *Tex Heart Inst J* 21:91-97; Graiger et al. (1993) *Cardiovasc Res* 27:2238-2247; and Grainger et al. (1993) *Biochem J* 294:109-112). Loss of sensitivity to TGFβ, or alternatively, the overriding of this inhibitory stimulus such as by PDGF autostimulation, can be a contributory factor to abnormal smooth muscle proliferation in restinosis. It may therefore be possible to treat or prevent restinosis by the use of gene therapy with gene constructs of the present invention which mimic induction by TGFβ. The *PDF* gene construct can be delivered, for example, by percutaneous transluminal gene transfer (Mazur et al. (1994) *Tex Heart Inst J* 21:104-111) using viral or liposomal delivery compositions. An exemplary adenovirus-mediated gene transfer technique and compositions for treatment of cardiac or vascular smooth muscle is provided in PCT publication WO 94/11506.

TGFβ's and wnt's also play a significant role in local glomerular and interstitial sites in human kidney development and disease. Consequently, the subject method provides a method of treating or inhibiting glomerulopathies and other renal proliferative disorders comprising the *in vivo* delivery of a subject *PDF* therapeutic.

Yet another aspect of the present invention concerns the therapeutic application of a *PDF* therapeutic to enhance survival of neurons and other neuronal cells in both the central nervous system and the peripheral nervous system. The ability of TGF-β factors

to regulate neuronal differentiation during development of the nervous system and also in the adult state indicates that certain of the *PDF* proteins can be reasonably expected to participate in control of adult neurons with regard to maintenance, functional performance, and aging of normal cells; repair and regeneration processes in chemically or mechanically lesioned cells; and prevention of degeneration and premature death which result from loss of differentiation in certain pathological conditions. In light of this understanding, the present invention specifically contemplates applications of the subject method to the treatment of (prevention and/or reduction of the severity of) neurological conditions deriving from: (i) acute, subacute, or chronic injury to the nervous system, including traumatic injury, chemical injury, vasal injury and deficits (such as the ischemia resulting from stroke), together with infectious/inflammatory and tumor-induced injury; (ii) aging of the nervous system including Alzheimer's disease; (iii) chronic neurodegenerative diseases of the nervous system, including Parkinson's disease, Huntington's chorea, amylotrophic lateral sclerosis and the like, as well as spinocerebellar degenerations; and (iv) chronic immunological diseases of the nervous system or affecting the nervous system, including multiple sclerosis.

Another aspect of the present invention comprises the inhibition of T cell activation. TGFβ is known to inhibit T cell proliferation and the *PDF*s described in the present invention could be used to ameliorate diseases that involve chronic inflammation. In addition, TGFβ has been associated with certain forms of tolerance (Chen et al. (1995) *Nature* 376:177-180) and the present invention could be used to induce T cell tolerance prior to receipt of an allo or xenograft or in cases of allergy or autoimmune disease.

In yet another embodiment, modulation of a *PDF1*-dependent pathway can be used to inhibit spermatogenesis. Spermatogenesis is a process involving mitotic replication of a pool of diploid stem cells, followed by meiosis and terminal differentiation of haploid cells into morphologically and functionally polarized spermatoza. This process exhibits both temporal and spatial regulation, as well as coordinated interaction between the germ and somatic cells. It has been previously shown that the signals mediated by the TGFβ superfamily, in particular activin, play significant roles in coupling such extracellular stimulus to regulation of mitotic, meiotic events which occur during spermatogenesis (Klaij, et al. (1994) *J. Endocrinol.* 141:131-141).

Likewise, members of the TGFβ family are important in the regulation of female reproductive organs (Wu, T.C. et al. (1994) *Mol. Reprod. Dev.* 38:9-15). Accordingly, TGFβ inhibitors, such as *PDF1* antagonists generated in the subject assays, may be

useful to prevent oocyte maturation as part of a contraceptive formulation. In other aspects, regulation of induction of meiotic maturation with *PDF1* therapeutics can be used synchronize oocyte populations for *in vitro* fertilization. Such a protocol can be used to provide a more homogeneous population of oocytes which are healthier and more viable and more prone to cleavage, fertilization and development to blastocyst stage. In addition the *PDF1* therapeutics could be used to treat other disorders of the female reproductive system which lead to infertility including polycysitic ovarian syndrome.

The *PDF* polypeptides of the present invention, or pharmaceutically acceptable salts thereof, may be conveniently formulated for administration with a biologically acceptable medium, such as water, buffered saline, polyol (for example, glycerol, propylene glycol, liquid polyethylene glycol and the like) or suitable mixtures thereof. The optimum concentration of the active ingredient(s) in the chosen medium can be determined empirically, according to procedures well known to medicinal chemists. As used herein, "biologically acceptable medium" includes any and all solvents, dispersion media, and the like which may be appropriate for the desired route of administration of the pharmaceutical preparation. The use of such media for pharmaceutically active substances is known in the art. Except insofar as any conventional media or agent is incompatible with the activity of the *PDF* polypeptide, its use in the pharmaceutical preparation of the invention is contemplated. Suitable vehicles and their formulation inclusive of other proteins are described, for example, in the book *Remmington's Pharmaceutical Sciences* (Remmington's Pharmaceutical Sciences. Mack Publishing Company, Easton, Pa., USA 1985). These vehicles include injectable "deposit formulations". Based on the above, such pharmaceutical formulations include, although not exclusively, solutions or freeze-dried powders of a PDF polypeptide in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered media at a suitable pH and isosmotic with physiological fluids. For illustrative purposes only and without being limited by the same, possible compositions or formulations which may be prepared in the form of solutions for the treatment of nervous system disorders with a PDF polypeptide are given in U.S. Patent No. 5,218,094. In the case of freeze-dried preparations, supporting excipients such as, but not exclusively, mannitol or glycine may be used and appropriate buffered solutions of the desired volume will be provided so as to obtain adequate isotonic buffered solutions of the desired pH. Similar solutions may also be used for the pharmaceutical compositions of *PDF* polypeptides in isotonic solutions of the desired volume and include, but not exclusively, the use of buffered saline solutions with phosphate or citrate at suitable concentrations so as to obtain at all times isotonic pharmaceutical preparations of the desired pH, (for example, neutral pH).

Methods of introduction of exogenous *PDF* polypeptides at the site of treatment include, but are not limited to, intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, oral, and intranasal. In addition, it may be desirable to introduce the pharmaceutical compositions of the invention into the central nervous system by any suitable route, including intraventricular and intrathecal injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

Methods of introduction may also be provided by rechargeable or biodegradable devices, particularly where gradients of *PDF* concentrations in a tissue is desired. Various slow release polymeric devices have been developed and tested *in vivo* in recent years for the controlled delivery of drugs, including proteinacious biopharmaceuticals. A variety of biocompatible polymers (including hydrogels), including both biodegradable and non-degradable polymers, can be used to form an implant for the sustained release of a PDF at a particular target site. Such embodiments of the present invention can be used for the delivery of an exogenously purified *PDF* polypeptides, which has been incorporated in the polymeric device, or for the delivery of *PDF* polypeptides produced by a cell encapsulated in the polymeric device. The generation of such implants is generally known in the art. See, for example, *Concise Encylopedia of Medical & Dental Materials*, ed. by David Williams (MIT Press: Cambridge, MA, 1990); the Sabel et al. U.S. Patent No. 4,883,666; Aebischer et al. U.S. Patent No. 4,892,538; Aebischer et al. U.S. Patent No. 5,106,627; Lim U.S. Patent No. 4,391,909; and Sefton U.S. Patent No. 4,353,888.

In yet another embodiment of the present invention, the pharmaceutical *PDF* polypeptide can be administered as part of a combinatorial therapy with other agents. For example, the combinatorial therapy can include a PDF protein with at least one trophic factor. Exemplary trophic factors include insulin like growth factor, nerve growth factor, cilliary neurotrophic growth factor, schwanoma-derived growth factor, glial growth factor, stiatal-derived neuronotrophic factor, platelet-derived growth factor, and scatter factor (HGF-SF). In another embodiment of the present invention, the PDF therapeutics are administered in conjunction with other agents, such as one or more of the BMP's, which may synergize with the PDF therapeutic and reduce the dosage required to achieve a beneficial effect.

Another aspect of the invention features transgenic non-human animals which express a heterologous *PDF* gene of the present invention, or which have had one or more genomic *PDF* genes disrupted in at least one of the tissue or cell-types of the animal. Accordingly, the invention features an animal model for developmental

diseases, which animal has *PDF* allele which is mis-expressed. For example, a mouse can be bred which has one or more *PDF* alleles deleted or otherwise rendered inactive. Such a mouse model can then be used to study disorders arising from mis-expressed *PDF* genes, as well as for evaluating potential therapies for similar disorders.

Another aspect of the present invention concerns transgenic animals which are comprised of cells (of that animal) which contain a transgene of the present invention and which preferably (though optionally) express an exogenous *PDF* protein in one or more cells in the animal. A *PDF* transgene can encode the wild-type form of the protein, or can encode homologs thereof, including both agonists and antagonists, as well as antisense constructs. In preferred embodiments, the expression of the transgene is restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, cis-acting sequences that control expression in the desired pattern. In the present invention, such mosaic expression of a *PDF* protein can be essential for many forms of lineage analysis and can additionally provide a means to assess the effects of, for example, lack of *PDF* expression which might grossly alter development in small patches of tissue within an otherwise normal embryo. Toward this and, tissue-specific regulatory sequences and conditional regulatory sequences can be used to control expression of the transgene in certain spatial patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences.

Genetic techniques which allow for the expression of transgenes can be regulated via site-specific genetic manipulation *in vivo* are known to those skilled in the art. For instance, genetic systems are available which allow for the regulated expression of a recombinase that catalyzes the genetic recombination a target sequence. As used herein, the phrase "target sequence" refers to a nucleotide sequence that is genetically recombined by a recombinase. The target sequence is flanked by recombinase recognition sequences and is generally either excised or inverted in cells expressing recombinase activity. Recombinase catalyzed recombination events can be designed such that recombination of the target sequence results in either the activation or repression of expression of one of the subject *PDF* proteins. For example, excision of a target sequence which interferes with the expression of a recombinant *PDF* gene, such as one which encodes an antagonistic homolog or an antisense transcript, can be designed to activate expression of that gene. This interference with expression of the protein can result from a variety of mechanisms, such as spatial separation of the *PDF* gene from the promoter element or an internal stop codon. Moreover, the transgene can be made wherein the coding sequence of the gene is flanked by recombinase recognition

sequences and is initially transfected into cells in a 3' to 5' orientation with respect to the promoter element. In such an instance, inversion of the target sequence will reorient the subject gene by placing the 5' end of the coding sequence in an orientation with respect to the promoter element which allow for promoter driven transcriptional activation.

In an illustrative embodiment, either the *cre/loxP* recombinase system of bacteriophage P1 (Lakso et al. (1992) *PNAS* 89:6232-6236; Orban et al. (1992) *PNAS* 89:6861-6865) or the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355; PCT publication WO 92/15694) can be used to generate *in vivo* site-specific genetic recombination systems. Cre recombinase catalyzes the site-specific recombination of an intervening target sequence located between *loxP* sequences. *loxP* sequences are 34 base pair nucleotide repeat sequences to which the Cre recombinase binds and are required for Cre recombinase mediated genetic recombination. The orientation of *loxP* sequences determines whether the intervening target sequence is excised or inverted when Cre recombinase is present (Abremski et al. (1984) *J. Biol. Chem.* 259:1509-1514); catalyzing the excision of the target sequence when the *loxP* sequences are oriented as direct repeats and catalyzes inversion of the target sequence when *loxP* sequences are oriented as inverted repeats.

Accordingly, genetic recombination of the target sequence is dependent on expression of the Cre recombinase. Expression of the recombinase can be regulated by promoter elements which are subject to regulatory control, e.g., tissue-specific, developmental stage-specific, inducible or repressible by externally added agents. This regulated control will result in genetic recombination of the target sequence only in cells where recombinase expression is mediated by the promoter element. Thus, the activation expression of a recombinant *PDF* protein can be regulated via control of recombinase expression.

Use of the *cre/loxP* recombinase system to regulate expression of a recombinant *PDF* protein requires the construction of a transgenic animal containing transgenes encoding both the Cre recombinase and the subject protein. Animals containing both the Cre recombinase and a recombinant *PDF* gene can be provided through the construction of "double" transgenic animals. A convenient method for providing such animals is to mate two transgenic animals each containing a transgene, e.g., a *PDF* gene and recombinase gene.

One advantage derived from initially constructing transgenic animals containing a *PDF* transgene in a recombinase-mediated expressible format derives from the likelihood that the subject protein, whether agonistic or antagonistic, can be deleterious

upon expression in the transgenic animal. In such an instance, a founder population, in which the subject transgene is silent in all tissues, can be propagated and maintained. Individuals of this founder population can be crossed with animals expressing the recombinase in, for example, one or more tissues and/or a desired temporal pattern. Thus, the creation of a founder population in which, for example, an antagonistic *PDF* transgene is silent will allow the study of progeny from that founder in which disruption of *PDF* mediated induction in a particular tissue or at certain developmental stages would result in, for example, a lethal phenotype.

Similar conditional transgenes can be provided using prokaryotic promoter sequences which require prokaryotic proteins to be simultaneous expressed in order to facilitate expression of the *PDF* transgene. Exemplary promoters and the corresponding trans-activating prokaryotic proteins are given in U.S. Patent No. 4,833,080.

Moreover, expression of the conditional transgenes can be induced by gene therapy-like methods wherein a gene encoding the trans-activating protein, e.g. a recombinase or a prokaryotic protein, is delivered to the tissue and caused to be expressed, such as in a cell-type specific manner. By this method, a *PDF* transgene could remain silent into adulthood until "turned on" by the introduction of the trans-activator.

In an exemplary embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonic target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonic target cell. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster et al. (1985) *PNAS* 82:4438-4442). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating transgenes in practicing the invention.

Retroviral infection can also be used to introduce *PDF* transgenes into a non-human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection

(Jaenich, R. (1976) *PNAS* 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (*Manipulating the Mouse Embryo*, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al. (1985) *PNAS* 82:6927-6931; Van der Putten et al. (1985) *PNAS* 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart et al. (1987) *EMBO J.* 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) *Nature* 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) *supra*).

A third type of target cell for transgene introduction is the embryonic stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (Evans et al. (1981) *Nature* 292:154-156; Bradley et al. (1984) *Nature* 309:255-258; Gossler et al. (1986) *PNAS* 83: 9065-9069; and Robertson et al. (1986) *Nature* 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) *Science* 240:1468-1474.

Methods of making *PDF* knock-out or disruption transgenic animals are also generally known. See, for example, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Recombinase dependent knockouts can also be generated, e.g. by homologous recombination to insert recombinase target sequences flanking portions of an endogenous *PDF* gene, such that tissue specific and/or temporal control of inactivation of a *PDF* allele can be controlled as above.

Exemplification

The invention, now being generally described, will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention and are not intended to limit the invention.

Example 1

RT-PCR Cloning of PDF cDNAs

RNA isolation and PCR amplification of PDF1 and PDF2 related genes from e13-e14-day fetal mouse pancreatic buds.

Poly(A)-RNA was prepared from pooled dorsal and ventral pancreatic buds isolated from e13- and e14-day mouse fetuses (40-50) using the QuickPrep Micro mRNA Purification Kit from Pharmacia Biotech. 1 ug poly(A)* was used in two first-strand cDNA reactions with either oligo(dT) or random hexanucleotides as primers for the reverse transcriptase reaction using the First-Strand cDNA Synthesis Kit from Pharmacia Biotech. One-third of each of the two cDNA reactions was combined and used as template for PCR amplification using 100 pmol of the following degenerate primers in a reaction of 50pl:

PDF1-primers from BMP:
5'-TGGAATTCTGG(ACG)A(ACGT)GA(CT)TGGAT(ACT)(AG)T(ACGT)GC-3'
5'-GAGGATCCA(AG)(ACGT)GT(CT)TG(ACGT)AC(AGT)AT(ACGT)GC(AG)TG-3'

PDF2-primers from Wnt:
5'-GGGGAATTCCA(AG)GA(AG)TG(CT)AA(CT)CAT-3'
5'-GAGGATCCA(AG)CA(AG)CACCA(AG)TG(AG)AA-3'

The degenerate positions in the above oligos are in parenthesis and restriction sites are underlined. The reaction was cycled twice between 94degreesC(50s), 50degreesC(2 min), and 72degreesC(2 min), followed by 28 rounds of 94degreesC(50s), 55degreesC(2 min), and 72degreesC(1.5min). The reaction products were purified, digested with BamHI and EcoRI (for BMP) or EcoRI and Xbal (for Wnt), size selected by agarose gel electrophoresis, and cloned into the pBluscript SKII+vector. Clones were picked randomly and analyzed by on a 7% sequencing gel by comparing their T ladders. One member of each class was sequenced completely using the T7 Sequencing Kit from Pharmacia Biotech.

cDNA isolation and sequencing.

An e13-e14 mouse dorsal and ventral pancreatic bud cDNA library of 10 independent clones was constructed in ZaPPII (Strategene) using 4ug of the poly(A)* RAN described above. After amplifying the library, 10 clones were screened under standard hybridization conditions, and a P random labeled probe was derived from the insert of pBluescript SKII+clone B6, representing *PDF1*, as well as from clone W27, representing the new *PDF2*. One (*PDF1*) and four *PDF2* positive clones were plaque purified and converted into pBluescript plasmids. Sequence analysis was performed by a combination of primer walking and the subcloning of small restriction fragments into pBluescript SKII+. The sequence within and adjacent to the long ORF was determined on both strands by the dideoxy chain termination method (Sangeret al., PNAS 74:5463-5467, 1977) using a T7 polymerase sequencing kit from Pharmacia Biotech.

In situ hybridization.

A Pst1 internal fragment representing coding sequences of *PDF1* and *PDF2*, respectively, was subcloned into pBluescript SKII+, linearized with Xhol (for both *PDF1* and *PDF2*) and used to generate a 410 bp (*PDF1*) and a 466 bp (*PDF2*) digoxygenim-11-UTP (Boeringer Mannheim, Germany) labeled antisense RNA probe using T3 RNA polymerase. Mouse embryos of different stages were fixed in 4% paraformaldehyde, and 10pm cryostat sections were mounted on SuperFrost/Plus slides (Menzel-Glaser, Germany). In situ hybridization was performed essentially as described by Schaeren-Wiemers and Gerfin-Moser (Histochemistry 100: 431-440, 1993).

All of the above-cited references and publications are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific polypeptides, nucleic acids, methods, assays and reagents described herein. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

What is claimed is:

1. An isolated or recombinant polypeptide comprising an amino acid sequence of a pancreatic-derived factor (*PDF*) protein.

2. An isolated or recombinant polypeptide comprising an amino acid sequence identical or at least 80% homologous to the pancreatic-derived factor (*PDF*) polypeptide designated by SEQ ID No. 2, 4 or 6, which polypeptide modulates at least one of growth, differentiation or paracrine secretion by a pancreatic cell, or a precursor thereof.

3. The polypeptide of claim 2, wherein said polypeptide stimulates intracellular signal transduction pathways mediated by a TGFβ receptor or a Frizzled receptor.

4. The polypeptide of claim 2, wherein said polypeptide antagonizes intracellular signal transduction pathways mediated by a TGFβ receptor or a Frizzled receptor.

5. The polypeptide of any of claims 1 or 2, comprising an amino acid sequence at least 85% homologous to the amino acid sequence of SEQ ID No. 2, 4 or 6.

6. The polypeptide of claim 5, comprising an amino acid sequence at least 95% homologous to the amino acid sequence of SEQ ID No. 2, 4 or 6.

7. The polypeptide of claim 5, comprising an amino acid sequence designated in SEQ ID No. 2, 4 or 6.

8. The polypeptide of any of claims 1 or 2, which polypeptide is encoded by a mammalian gene.

9. The polypeptide of claim 8, wherein the mammalian gene is a human gene.

10. The polypeptide of any of claims 1 or 2, comprising an amino acid sequence encoded by a nucleic acid which hybridizes under stringent conditions to the coding sequence designated in SEQ ID No. 1, 3 or 5.

11. The polypeptide of any of claims 1 or 2, which polypeptide is a fusion protein further comprising, a second polypeptide portion having an amino acid sequence from a protein different than said PDF polypeptide.

12. The polypeptide of any of claims 1, 2 or 3, which polypeptide comprises at least 50 contiguous amino acid residues of SEQ ID No. 2, 4 or 6, respectively.

13. The polypeptide of any of claims 1, 2 or 3, which polypeptide comprises at least 100 contiguous amino acid residues of SEQ ID No. 2 , 4 or 6, respectively.

14. The polypeptide of any of claims 1 or 2, which polypeptide is substantially free of other cellular proteins with each it naturally associates.

15. An isolated and/or recombinant *PDF* polypeptide of mammalian origin.

16. An isolated and/or recombinant *PDF* polypeptide of human origin.

17. An isolated and/or recombinant polypeptide comprising an amino acid sequence cross-reactive with an antibody specific for a mammalian *PDF* protein designated in SEQ ID No. 2, 4 or 6, which polypeptide modulates at least one of growth, differentiation or paracrine secretion by a pancreatic cell, or a precursor thereof.

18. A purified or recombinant *PDF* polypeptide comprising a core motif CKKTPLTIDFKEIG.

19. The *PDF* polypeptide of claim 18, wherein said polypeptide modulates intracellular signal transduction pathways mediated by a TGFβ receptor.

20. A nucleic acid which encodes a *PDF* polypeptide of mammalian origin.

21. A purified or recombinant *PDF* polypeptide encoded by a nucleic acid which hybridizes under stringent conditions to a nucleotide sequence designated in SEQ ID No. 1, 3 or 5.

22. An isolated nucleic acid encoding a polypeptide including a core motif represented by CKKTPLTIDFKEIG, which polypeptide specifically modulates the signal transduction activity of a receptor for a transforming growth factor β (TGFβ).

23. The nucleic acid of claim 22, wherein said polypeptide comprises the amino acid sequence represented by SEQ ID No. 2 or 6.

24. The nucleic acid of claim 22, wherein said polypeptide is a fusion protein further comprising, in addition to said core motif, a second polypeptide sequence having an amino acid sequence unrelated to a nucleic acid sequence.

25. The nucleic acid of claim 22, which nucleic acid hybridizes under stringent conditions to a nucleic acid probe having a sequence represented by at least 60 consecutive nucleotides of sense or antisense of SEQ ID No. 1 or 5.

26. The nucleic acid of any of claims 20, 21 or 22, further comprising a transcriptional regulatory sequence operably linked to said nucleotide sequence so as to render said nucleic acid suitable for use as an expression vector.

27. An expression vector, capable of replicating in at least one of a prokaryotic cell and eukaryotic cell, comprising the nucleic acid of claim 26.

28. A host cell transfected with the expression vector of claim 27 and expressing said recombinant polypeptide.

29. A method of producing a recombinant *PDF* polypeptide comprising culturing the cell of claim 28 in a cell culture medium to express said recombinant polypeptide and isolating said recombinant polypeptide from said cell culture.

30. A transgenic animal having cells which harbor a transgene encoding a *PDF* polypeptide, which animals are vertebrates.

31. A transgenic animal having cells in which a gene for a *PDF* is disrupted, which animals are vertebrates.

32. A recombinant transfection system, comprising
    (i) a gene construct including the nucleic acid of claim 26 and operably linked to a transcriptional regulatory sequence for causing expression of said *PDF* polypeptide in eukaryotic cells, and
    (ii) a gene delivery composition for delivering said gene construct to a cell and causing the cell to be transfected with said gene construct.

33. The recombinant transfection system of claim 32, wherein the gene delivery composition is selected from a group consisting of a recombinant viral particle, a liposome, and a poly-cationic nucleic acid binding agent.

34. A nucleic acid composition comprising a substantially purified oligonucleotide, said oligonucleotide including a region of nucleotide sequence which hybridizes under stringent conditions to at least 25 consecutive nucleotides of sense or antisense sequence of a mammalian *PDF* gene.

35. The nucleic acid composition of claim 34, which oligonucleotide hybridizes under stringent conditions to at least 50 consecutive nucleotides of sense or antisense sequence of a *PDF* gene.

36. The nucleic acid composition of claim 34, which oligonucleotide hybridizes under stringent conditions to the sense or antisense nucleotide sequence designated in SEQ ID No. 1, 3 or 5.

37. The nucleic acid composition of claim 34, wherein said oligonucleotide further comprises a label group attached thereto and able to be detected.

38. The nucleic acid composition of claim 34, wherein said oligonucleotide has at least one non-hydrolyzable bond between two adjacent nucleotide subunits.

39. A test kit for detecting cells which contain a *PDF* mRNA transcript, comprising the nucleic acid composition of claim 34 for measuring, in a sample of cells, a level of nucleic acid encoding a *PDF* protein.

40. A method for modulating one or more of growth, differentiation, or survival of a mammalian cell responsive to *PDF*-mediated induction, comprising treating the cell with an effective amount of an agent which modulates the signal transduction activity of a *PDF* polypeptide thereby altering, relative to the cell in the absence of the agent, at least one of (i) rate of growth, (ii) differentiation, or (iii) survival of the cell.

41. The method of claim 40, wherein said agent mimics the effects of a naturally-occurring *PDF* protein on said cell.

42. The method of claim 40, wherein said agent antagonizes the effects of a naturally-occurring *PDF* protein on said cell.

43. The method of claim 40, wherein the cell is a pancreatic cell, and the agent modulates carbohydrate metabolism.

44. An antibody to a *PDF* polypeptide.

45. The antibody of claim 5344 wherein said antibody is monoclonal.

46. A *PDF* polypeptide which specifically modulates the signal transduction activity of a TGFβ receptor other than a TGFβ receptor for a BMP-2.

47. A diagnostic assay for identifying a cell or cells at risk for a disorder characterized by unwanted cell proliferation or differentiation, comprising detecting, in a cell sample, the presence or absence of a genetic lesion characterized by at least one of (i) aberrant modification or mutation of a gene encoding a *PDF* protein, and (ii) mis-expression of said gene; wherein a wild-type form of said gene encodes a *PDF* protein characterized by an ability to modulate the signal transduction activity of a TGFb receptor.

48. The assay of claim 47, wherein detecting said lesion includes:
    i. providing a diagnostic probe comprising a nucleic acid including a region of nucleotide sequence which hybridizes to a sense or antisense sequence of said gene, or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with said gene;
    ii. combining said probe with nucleic acid of said cell sample; and
    iii. detecting, by hybridization of said probe to said cellular nucleic acid, the existence of at least one of a deletion of one or more nucleotides from said gene, an addition of one or more nucleotides to said gene, a substitution of one or more nucleotides of said gene, a gross chromosomal rearrangement of all or a portion of said gene, a gross alteration in the level of an mRNA transcript of said gene, or a non-wild type splicing pattern of an mRNA transcript of said gene.

49. The assay of claim 48, wherein hybridization of said probe further comprises subjecting the probe and cellular nucleic acid to a polymerase chain reaction (PCR) and detecting abnormalities in an amplified product.

50. The assay of claim 48, wherein hybridization of said probe further comprises subjecting the probe and cellular nucleic acid to a ligation chain reaction (LCR) and detecting abnormalities in an amplified product.

51. The assay of claim 48, wherein said probe hybridizes under stringent conditions to a nucleic acid designated by SEQ ID No. 1, 3 or 5.

52. The use of a pancreatic-derived factor (PDF) of claim 20, 21 or 22, in the manufacture of a medicament for the treatment of a pancreatic disorder or to modulate growth and/or differentiation of pancreatic cells or stem cells capable of differentiating to pancreatic cells.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: ONTOGENY, INC.

(ii) TITLE OF INVENTION: PANCREATIC-DERIVED FACTORS, AND USES RELATED THERETO

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Foley, Hoag & Eliot
(B) STREET: One Post Office Square
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02109

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Vincent, Matthew P.
(B) REGISTRATION NUMBER: 36,709
(C) REFERENCE/DOCKET NUMBER: ONV-022.25

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)832-1000
(B) TELEFAX: (617)832-7000

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..1317

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TT TTT GAG GTA CTA GAG AGT GCA GAC GGT AGC GAG GAG GAG AGG AGC      47
   Phe Glu Val Leu Glu Ser Ala Asp Gly Ser Glu Glu Glu Arg Ser
     1               5                  10                  15

ATG CTG GTC TTA GTA TCA ACA GAG ATC TAC GGA ACC AAC AGT GAG TGG     95
Met Leu Val Leu Val Ser Thr Glu Ile Tyr Gly Thr Asn Ser Glu Trp
                 20                  25                  30

GAG ACA TTT GAC GTC ACA GAT GCC ACC AGA CGT TGG CAA AAG TCA GGC    143
Glu Thr Phe Asp Val Thr Asp Ala Thr Arg Arg Trp Gln Lys Ser Gly
             35                  40                  45

CCA TCA ACC CAT CAG CTG GAG ATC CAC ATA GAA AGC AGA CAA AAC CAA    191
Pro Ser Thr His Gln Leu Glu Ile His Ile Glu Ser Arg Gln Asn Gln
         50                  55                  60

GCT GAG GAC ACC GGA AGG GGA CAA CTG GAA ATA GAT ATG AGT GCC CAA    239
Ala Glu Asp Thr Gly Arg Gly Gln Leu Glu Ile Asp Met Ser Ala Gln
     65                  70                  75

AAT AAG CAT GAC CCT TTG CTG GTT GTG TTT TCT GAT GAC CAA AGC AAT    287
Asn Lys His Asp Pro Leu Leu Val Val Phe Ser Asp Asp Gln Ser Asn
 80                  85                  90                  95

GAC AAG GAG CAG AAA GAA GAA CTG AAC GAA TTG ATC ACC CAT GAG CAG    335
Asp Lys Glu Gln Lys Glu Glu Leu Asn Glu Leu Ile Thr His Glu Gln
                100                 105                 110

GAT CTG GAC CTG GAC TCA GAT GCT TTC TTC AGT GGG CCC GAT GAA GAG    383
Asp Leu Asp Leu Asp Ser Asp Ala Phe Phe Ser Gly Pro Asp Glu Glu
            115                 120                 125

GCT CTG CTG CAG ATG AGG TCG AAC ATG ATT GAT GAT TCG TCC GCT CGG    431
Ala Leu Leu Gln Met Arg Ser Asn Met Ile Asp Asp Ser Ser Ala Arg
        130                 135                 140

ATC AGG AGG AAC GCC AAG GGG AAC TAC TGT AAG AAG ACC CCA CTA TAC    479
Ile Arg Arg Asn Ala Lys Gly Asn Tyr Cys Lys Lys Thr Pro Leu Tyr
    145                 150                 155

ATC GAC TTC AAG GAG ATT GGG TGG GAC TCC TGG ATC ATC GCT CCT CCT    527
Ile Asp Phe Lys Glu Ile Gly Trp Asp Ser Trp Ile Ile Ala Pro Pro
160                 165                 170                 175

GGG TAT GAA GCC TAT GAG TGC CGG GGT GTG TGT AAC TAC CCT CTG GCG    575
Gly Tyr Glu Ala Tyr Glu Cys Arg Gly Val Cys Asn Tyr Pro Leu Ala
                180                 185                 190

GAG ACC TCA CAC CTC AAA ACA CGC AAT TAT TCC GGC CTG GTC CAC CTC    623
Glu Thr Ser His Leu Lys Thr Arg Asn Tyr Ser Gly Leu Val His Leu
            195                 200                 205

AAG AAT TCC CAG AAA GCT TCC AAA GCC TGC TGT GTG CAT AAG CAT GAC    671
```

```
Lys Asn Ser Gln Lys Ala Ser Lys Ala Cys Cys Val His Lys His Asp
        210                 215                 220

CCT TTG CTG GTT GTG TTT TCT GAT GAC CAA AGC AAT GAC AAG GAG CAG      719
Pro Leu Leu Val Val Phe Ser Asp Asp Gln Ser Asn Asp Lys Glu Gln
    225                 230                 235

AAA GAA GAA CTG AAC GAA TTG ATC ACC CAT GAG CAG GAT CTG GAC CTG      767
Lys Glu Glu Leu Asn Glu Leu Ile Thr His Glu Gln Asp Leu Asp Leu
240                 245                 250                 255

GAC TCA GAT GCT TTC TTC AGT GGG CCC GAT GAA GAG GCT CTG CTG CAG      815
Asp Ser Asp Ala Phe Phe Ser Gly Pro Asp Glu Glu Ala Leu Leu Gln
                260                 265                 270

ATG AGG TCG AAC ATG ATT GAT GAT TCG TCC GCT CGG ATC AGG AGG AAC      863
Met Arg Ser Asn Met Ile Asp Asp Ser Ser Ala Arg Ile Arg Arg Asn
            275                 280                 285

GCC AAG GGG AAC TAC TGT AAG AAG ACC CCA CTA TAC ATC GAC TTC AAG      911
Ala Lys Gly Asn Tyr Cys Lys Lys Thr Pro Leu Tyr Ile Asp Phe Lys
        290                 295                 300

GAG ATT GGG TGG GAC TCC TGG ATC ATC GCT CCT CCT GGG TAT GAA GCC      959
Glu Ile Gly Trp Asp Ser Trp Ile Ile Ala Pro Pro Gly Tyr Glu Ala
    305                 310                 315

TAT GAG TGC CGG GGT GTG TGT AAC TAC CCT CTG GCG GAG ACC TCA CAC     1007
Tyr Glu Cys Arg Gly Val Cys Asn Tyr Pro Leu Ala Glu Thr Ser His
320                 325                 330                 335

CTC AAA ACA CGC AAT TAT TCC GGC CTG GTC CAC CTC AAG AAT TCC CAG     1055
Leu Lys Thr Arg Asn Tyr Ser Gly Leu Val His Leu Lys Asn Ser Gln
                340                 345                 350

AAA GCT TCC AAA GCC TGC TGT GTG CCC ACG AAG CTG GAT CCC ATC TCC     1103
Lys Ala Ser Lys Ala Cys Cys Val Pro Thr Lys Leu Asp Pro Ile Ser
            355                 360                 365

ATC CTC TAT TTA GAT AAA GGT GTC GTC ACC TAC AAG TTT AAA TAT GAA     1151
Ile Leu Tyr Leu Asp Lys Gly Val Val Thr Tyr Lys Phe Lys Tyr Glu
        370                 375                 380

GGG ATG GCT GTG TCT GAG TGT GGC TGT AGA TAG GAG AGG AGA GGC GTC     1199
Gly Met Ala Val Ser Glu Cys Gly Cys Arg  *  Glu Arg Arg Gly Val
    385                 390                 395

CCA TGG CTT ATT TAA TAA CTG CAG ATT TAA TGA AGG TGT ACA GAT AAT     1247
Pro Trp Leu Ile  *   *  Leu Gln Ile  *   *  Arg Cys Thr Asp Asn
400                 405                 410                 415

AGA GGT GCC ACC TTA GGA CTG TGC TGT AGG AAT GTG TAT TTG CTC ACA     1295
Arg Gly Ala Thr Leu Gly Leu Cys Cys Arg Asn Val Tyr Leu Leu Thr
                420                 425                 430

TCA GTC CAT CAG TAC TCT CAT A                                       1317
```

Ser Val His Gln Tyr Ser His
435

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 438 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Phe Glu Val Leu Glu Ser Ala Asp Gly Ser Glu Glu Glu Arg Ser Met
1 5 10 15

Leu Val Leu Val Ser Thr Glu Ile Tyr Gly Thr Asn Ser Glu Trp Glu
20 25 30

Thr Phe Asp Val Thr Asp Ala Thr Arg Arg Trp Gln Lys Ser Gly Pro
35 40 45

Ser Thr His Gln Leu Glu Ile His Ile Glu Ser Arg Gln Asn Gln Ala
50 55 60

Glu Asp Thr Gly Arg Gly Gln Leu Glu Ile Asp Met Ser Ala Gln Asn
65 70 75 80

Lys His Asp Pro Leu Leu Val Val Phe Ser Asp Asp Gln Ser Asn Asp
85 90 95

Lys Glu Gln Lys Glu Glu Leu Asn Glu Leu Ile Thr His Glu Gln Asp
100 105 110

Leu Asp Leu Asp Ser Asp Ala Phe Phe Ser Gly Pro Asp Glu Glu Ala
115 120 125

Leu Leu Gln Met Arg Ser Asn Met Ile Asp Asp Ser Ser Ala Arg Ile
130 135 140

Arg Arg Asn Ala Lys Gly Asn Tyr Cys Lys Lys Thr Pro Leu Tyr Ile
145 150 155 160

Asp Phe Lys Glu Ile Gly Trp Asp Ser Trp Ile Ile Ala Pro Pro Gly
165 170 175

Tyr Glu Ala Tyr Glu Cys Arg Gly Val Cys Asn Tyr Pro Leu Ala Glu
180 185 190

Thr Ser His Leu Lys Thr Arg Asn Tyr Ser Gly Leu Val His Leu Lys
195 200 205

Asn Ser Gln Lys Ala Ser Lys Ala Cys Cys Val His Lys His Asp Pro
210 215 220

```
Leu Leu Val Val Phe Ser Asp Asp Gln Ser Asn Asp Lys Glu Gln Lys
225                 230                 235                 240

Glu Glu Leu Asn Glu Leu Ile Thr His Glu Gln Asp Leu Asp Leu Asp
                245                 250                 255

Ser Asp Ala Phe Phe Ser Gly Pro Asp Glu Glu Ala Leu Leu Gln Met
            260                 265                 270

Arg Ser Asn Met Ile Asp Asp Ser Ser Ala Arg Ile Arg Arg Asn Ala
        275                 280                 285

Lys Gly Asn Tyr Cys Lys Lys Thr Pro Leu Tyr Ile Asp Phe Lys Glu
    290                 295                 300

Ile Gly Trp Asp Ser Trp Ile Ile Ala Pro Pro Gly Tyr Glu Ala Tyr
305                 310                 315                 320

Glu Cys Arg Gly Val Cys Asn Tyr Pro Leu Ala Glu Thr Ser His Leu
                325                 330                 335

Lys Thr Arg Asn Tyr Ser Gly Leu Val His Leu Lys Asn Ser Gln Lys
            340                 345                 350

Ala Ser Lys Ala Cys Cys Val Pro Thr Lys Leu Asp Pro Ile Ser Ile
        355                 360                 365

Leu Tyr Leu Asp Lys Gly Val Val Thr Tyr Lys Phe Lys Tyr Glu Gly
    370                 375                 380

Met Ala Val Ser Glu Cys Gly Cys Arg  *  Glu Arg Arg Gly Val Pro
385                 390                 395                 400

Trp Leu Ile  *   *  Leu Gln Ile  *   *  Arg Cys Thr Asp Asn Arg
                405                 410                 415

Gly Ala Thr Leu Gly Leu Cys Cys Arg Asn Val Tyr Leu Leu Thr Ser
            420                 425                 430

Val His Gln Tyr Ser His
        435
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 645 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS

(B) LOCATION: 3..645

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CC CGG TTC CAT CTT CGC GCT GGT CTT GCC TGC CTG CTG CTG CTG CTA         47
   Arg Phe His Leu Arg Ala Gly Leu Ala Cys Leu Leu Leu Leu Leu
    1               5                  10                  15

CTC CTG ACT CTG CCG GCC CGT GTA GAC ACG TCC TGG TGG TAC ATA GGG        95
Leu Leu Thr Leu Pro Ala Arg Val Asp Thr Ser Trp Trp Tyr Ile Gly
                20                  25                  30

GCT CTG GGA GCC CGA GTG ATC TGT GAC AAC ATC CCC GGT CTG GTG AGC       143
Ala Leu Gly Ala Arg Val Ile Cys Asp Asn Ile Pro Gly Leu Val Ser
            35                  40                  45

CGG CAG CGG CAG TTG TGT CAA CGC TAC CCA GAC ATC ATG CGC TCA GTA       191
Arg Gln Arg Gln Leu Cys Gln Arg Tyr Pro Asp Ile Met Arg Ser Val
        50                  55                  60

GGT GAG GGT GCC CGG GAA TGG ATC CGA GAG TGC CAG CAC CAG TTC CGT       239
Gly Glu Gly Ala Arg Glu Trp Ile Arg Glu Cys Gln His Gln Phe Arg
    65                  70                  75

GAC CAC CGC TGG AAT TGC ACC ACA CTG GAC CGG GAC CAC ACT GTC TTT       287
Asp His Arg Trp Asn Cys Thr Thr Leu Asp Arg Asp His Thr Val Phe
80                  85                  90                  95

GGC CGC GCC ATG CTC AGA AGC AGC CGG GAC GGA GCG TTC GTC TAT GCT       335
Gly Arg Ala Met Leu Arg Ser Ser Arg Asp Gly Ala Phe Val Tyr Ala
                100                 105                 110

ATC TCG TCA GCA GGA GTG GTC CTC GCT ATC ACT CGG GCC TGC AGC CAG       383
Ile Ser Ser Ala Gly Val Val Leu Ala Ile Thr Arg Ala Cys Ser Gln
            115                 120                 125

GGT GAG CTG AGC GTG TGC AGC TGT GAC CAT ATA CCC CGC GGT CGG CAC       431
Gly Glu Leu Ser Val Cys Ser Cys Asp His Ile Pro Arg Gly Arg His
        130                 135                 140

CAT GAT CAA CGA GGG GAC TTT GAC TGG GGT GGC TGT AGT GAC AAC ATC       479
His Asp Gln Arg Gly Asp Phe Asp Trp Gly Gly Cys Ser Asp Asn Ile
    145                 150                 155

CAT TAC GGT GTT CGC TTT GCC AAG GCT TTT GTG GAT GCC AAA GAG AAG       527
His Tyr Gly Val Arg Phe Ala Lys Ala Phe Val Asp Ala Lys Glu Lys
160                 165                 170                 175

AGG CTT AAG GAT GCC CGG GCC CTC ATG AAC TTA CAC AAC AAC CGC TGT       575
Arg Leu Lys Asp Ala Arg Ala Leu Met Asn Leu His Asn Asn Arg Cys
                180                 185                 190

GGT CGC ACG GCT GTT CGG AGA TTC CTG AAG CTG GAG TGC AAG TGT CAC       623
Gly Arg Thr Ala Val Arg Arg Phe Leu Lys Leu Glu Cys Lys Cys His
            195                 200                 205
```

GGT GTG AGT GGC TCC TGT ACT C 645
Gly Val Ser Gly Ser Cys Thr
210

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Arg Phe His Leu Arg Ala Gly Leu Ala Cys Leu Leu Leu Leu Leu Leu
1 5 10 15

Leu Thr Leu Pro Ala Arg Val Asp Thr Ser Trp Trp Tyr Ile Gly Ala
20 25 30

Leu Gly Ala Arg Val Ile Cys Asp Asn Ile Pro Gly Leu Val Ser Arg
35 40 45

Gln Arg Gln Leu Cys Gln Arg Tyr Pro Asp Ile Met Arg Ser Val Gly
50 55 60

Glu Gly Ala Arg Glu Trp Ile Arg Glu Cys Gln His Gln Phe Arg Asp
65 70 75 80

His Arg Trp Asn Cys Thr Thr Leu Asp Arg Asp His Thr Val Phe Gly
85 90 95

Arg Ala Met Leu Arg Ser Ser Arg Asp Gly Ala Phe Val Tyr Ala Ile
100 105 110

Ser Ser Ala Gly Val Val Leu Ala Ile Thr Arg Ala Cys Ser Gln Gly
115 120 125

Glu Leu Ser Val Cys Ser Cys Asp His Ile Pro Arg Gly Arg His His
130 135 140

Asp Gln Arg Gly Asp Phe Asp Trp Gly Gly Cys Ser Asp Asn Ile His
145 150 155 160

Tyr Gly Val Arg Phe Ala Lys Ala Phe Val Asp Ala Lys Glu Lys Arg
165 170 175

Leu Lys Asp Ala Arg Ala Leu Met Asn Leu His Asn Asn Arg Cys Gly
180 185 190

Arg Thr Ala Val Arg Arg Phe Leu Lys Leu Glu Cys Lys Cys His Gly
195 200 205

```
Val Ser Gly Ser Cys Thr
    210
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1522 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 124..1386

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GGGATTTTGG CCGACTTTGC CCTTTTGGCA GCCCTTTTCG CGGGGACCCA CCAGAGCCCT    60

GGTAATGGTA GCGACCGGCG CTCAGCTGGA ATTCGCGGCC GCGTCGACGC TGAGCAGAGA   120

GTC ATG GGG TCT CTG GTT CTG CCG CTG AGC GCC GTC TTC TGC CTG GTG     168
    Met Gly Ser Leu Val Leu Pro Leu Ser Ala Val Phe Cys Leu Val
      1               5                  10                  15

GCT CAC TCG GCT TCT GGC AGC CCC ATT ATG GGC CTT GAG CAG TCG CCC     216
Ala His Ser Ala Ser Gly Ser Pro Ile Met Gly Leu Glu Gln Ser Pro
                 20                  25                  30

CTG GAA GAA GAC ATG CCC TTC TTT GAT GAT ATC TTC ACG GAG CAA GAT     264
Leu Glu Glu Asp Met Pro Phe Phe Asp Asp Ile Phe Thr Glu Gln Asp
             35                  40                  45

GGT ATT GAC TTC AAC ACA CTG CTG CAG AGC ATG AAG AAC GAG TTT CTC     312
Gly Ile Asp Phe Asn Thr Leu Leu Gln Ser Met Lys Asn Glu Phe Leu
         50                  55                  60

AAG ACG CTG AAC TTG TCG GAC ATT CCT GTG CAG GAC ACG GGC AGA GTG     360
Lys Thr Leu Asn Leu Ser Asp Ile Pro Val Gln Asp Thr Gly Arg Val
     65                  70                  75

GAT CCA CCA GAG TAC ATG CTG GAG CTC TAC AAC AAA TTC GCC ACA GAC     408
Asp Pro Pro Glu Tyr Met Leu Glu Leu Tyr Asn Lys Phe Ala Thr Asp
 80                  85                  90                  95

CGG ACC TCC ATG CCG TCT GCT AAC ATC ATC CGG AGC TTC AAG AAC GAA     456
Arg Thr Ser Met Pro Ser Ala Asn Ile Ile Arg Ser Phe Lys Asn Glu
                100                 105                 110

GAT CTG TTT TCT CAA CCA GTC ACT TTT AAT GGG CTC CGG AAA TAT CCT     504
Asp Leu Phe Ser Gln Pro Val Thr Phe Asn Gly Leu Arg Lys Tyr Pro
```

115 120 125

CTC CTC TTC AAT GTG TCT ATC CCT CAC CAT GAA GAG GTC GTC ATG GCT 552
Leu Leu Phe Asn Val Ser Ile Pro His His Glu Glu Val Val Met Ala
130 135 140

GAA CTG CGG TTG TAC ACG CTG GTG CAG AGA GAT CGC ATG ATG TAT GAT 600
Glu Leu Arg Leu Tyr Thr Leu Val Gln Arg Asp Arg Met Met Tyr Asp
145 150 155

GGC GTG GAC CGT AAA ATT ACC ATT TTT GAG GTA CTA GAG AGT GCA GAC 648
Gly Val Asp Arg Lys Ile Thr Ile Phe Glu Val Leu Glu Ser Ala Asp
160 165 170 175

GGT AGC GAG GAG GAG AGG AGC ATG CTG GTC TTA GTA TCA ACA GAG ATC 696
Gly Ser Glu Glu Glu Arg Ser Met Leu Val Leu Val Ser Thr Glu Ile
180 185 190

TAC GGA ACC AAC AGT GAG TGG GAG ACA TTT GAC GTC ACA GAT GCC ACC 744
Tyr Gly Thr Asn Ser Glu Trp Glu Thr Phe Asp Val Thr Asp Ala Thr
195 200 205

AGA CGT TGG CAA AAG TCA GGC CCA TCA ACC CAT CAG CTG GAG ATC CAC 792
Arg Arg Trp Gln Lys Ser Gly Pro Ser Thr His Gln Leu Glu Ile His
210 215 220

ATA GAA AGC AGA CAA AAC CAA GCT GAG GAC ACC GGA AGG GGA CAA CTG 840
Ile Glu Ser Arg Gln Asn Gln Ala Glu Asp Thr Gly Arg Gly Gln Leu
225 230 235

GAA ATA GAT ATG AGT GCC CAA AAT AAG CAT GAC CCT TTG CTG GTT GTG 888
Glu Ile Asp Met Ser Ala Gln Asn Lys His Asp Pro Leu Leu Val Val
240 245 250 255

TTT TCT GAT GAC CAA AGC AAT GAC AAG GAG CAG AAA GAA GAA CTG AAC 936
Phe Ser Asp Asp Gln Ser Asn Asp Lys Glu Gln Lys Glu Glu Leu Asn
260 265 270

GAA TTG ATC ACC CAT GAG CAG GAT CTG GAC CTG GAC TCA GAT GCT TTC 984
Glu Leu Ile Thr His Glu Gln Asp Leu Asp Leu Asp Ser Asp Ala Phe
275 280 285

TTC AGT GGG CCC GAT GAA GAG GCT CTG CTG CAG ATG AGG TCG AAC ATG 1032
Phe Ser Gly Pro Asp Glu Glu Ala Leu Leu Gln Met Arg Ser Asn Met
290 295 300

ATT GAT GAT TCG TCC GCT CGG ATC AGG AGG AAC GCC AAG GGG AAC TAC 1080
Ile Asp Asp Ser Ser Ala Arg Ile Arg Arg Asn Ala Lys Gly Asn Tyr
305 310 315

TGT AAG AAG ACC CCA CTA TAC ATC GAC TTC AAG GAG ATT GGG TGG GAC 1128
Cys Lys Lys Thr Pro Leu Tyr Ile Asp Phe Lys Glu Ile Gly Trp Asp
320 325 330 335

TCC TGG ATC ATC GCT CCT CCT GGG TAT GAA GCC TAT GAG TGC CGG GGT 1176
Ser Trp Ile Ile Ala Pro Pro Gly Tyr Glu Ala Tyr Glu Cys Arg Gly

340 345 350

GTG TGT AAC TAC CCT CTG GCG GAG CAC CTC ACA CCT ACA AAA CAC GCA 1224
Val Cys Asn Tyr Pro Leu Ala Glu His Leu Thr Pro Thr Lys His Ala
355 360 365

ATT ATT CAG GCC TTG GTC CAC CTC AAG AAT TCC CAG AAA GCT TCC AAA 1272
Ile Ile Gln Ala Leu Val His Leu Lys Asn Ser Gln Lys Ala Ser Lys
370 375 380

GCC TGC TGT GTG CCC ACG AAG CTG GAT CCC ATC TCC ATC CTC TAT TTA 1320
Ala Cys Cys Val Pro Thr Lys Leu Asp Pro Ile Ser Ile Leu Tyr Leu
385 390 395

GAT AAA GGT GTC GTC ACC TAC AAG TTT AAA TAT GAA GGG ATG GCT GTG 1368
Asp Lys Gly Val Val Thr Tyr Lys Phe Lys Tyr Glu Gly Met Ala Val
400 405 410 415

TCT GAG TGT GGC TGT AGA TAGGAGAGGA GAGGCGTCCC ATGGCTTATT 1416
Ser Glu Cys Gly Cys Arg
420

TAATAACTGC AGATTTAATG AAGGTGTACA GATAATAGAG GTGCCACCTT AGGACTGTGC 1476

TGTAGGAATG TGTATTTGCT CACATCAGTC CATCAGTACT CTCATA 1522

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 421 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gly Ser Leu Val Leu Pro Leu Ser Ala Val Phe Cys Leu Val Ala
1 5 10 15

His Ser Ala Ser Gly Ser Pro Ile Met Gly Leu Glu Gln Ser Pro Leu
20 25 30

Glu Glu Asp Met Pro Phe Phe Asp Asp Ile Phe Thr Glu Gln Asp Gly
35 40 45

Ile Asp Phe Asn Thr Leu Leu Gln Ser Met Lys Asn Glu Phe Leu Lys
50 55 60

Thr Leu Asn Leu Ser Asp Ile Pro Val Gln Asp Thr Gly Arg Val Asp
65 70 75 80

Pro Pro Glu Tyr Met Leu Glu Leu Tyr Asn Lys Phe Ala Thr Asp Arg
85 90 95